

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 04:56:29 ; Search time 74.1768 Seconds  
(without alignments)  
1652.628 Million cell updates/sec

Title: US-10-617-038-21  
1110

Perfect score: 1 MNSHKPPSIWVGIDSKPA.....GNAVLOEAGCTLLVVGQOYL 279  
Sequence:

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21:\*

1:	geneseqp1980s:*
2:	geneseqp1990s:*
3:	geneseqp2000s:*
4:	geneseqp2001s:*
5:	geneseqp2002s:*
6:	geneseqp2003as:*
7:	geneseqp2003bs:*
8:	geneseqp2004s:*
9:	geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1410	100.0	279	8	AD137300 M. tuberc
2	431.5	30.6	268	8	AD137321 M. tuberc
3	416	29.5	272	8	AD137306 M. tuberc
4	257	18.2	295	8	AD137298 M. tuberc
5	214	15.2	297	3	AAV87878 M. tuberc
6	214	15.2	297	5	ABG30907 Mycobacte
7	207	14.7	301	4	AAV71872 C. glutam
8	207	14.7	301	4	AAV71865 C. glutam
9	207	14.7	301	4	AAV71865 C. glutam
10	207	14.7	301	4	AAV71865 C. glutam
11	207	14.7	301	5	ABG80324 C. glutam
12	207	14.7	301	5	ABG80324 C. glutam
13	194	13.8	316	4	AAV62080 Propionib
14	194	13.8	316	6	ABM58599 Propionib
15	180	12.8	300	6	ABU25937 Protein e
16	127.5	9.0	3227	8	ADQ91710 Polypeptid
17	127.5	9.0	3227	8	ABE87002 Streptomy
18	127.5	9.0	3227	8	ABE86799 Streptomy
19	118.5	8.4	2539	6	AAH33731 Protein e
20	115	8.2	3562	4	AAH82213 Polypeptid
21	113	8.0	7068	4	AAE10142 Streptomy
22	113	8.0	9477	4	AAE10144 Streptomy
23	112	7.9	11096	4	AAE10129 Streptomy
24	109.5	7.8	10917	6	AAE36132 Streptomy

25	107.5	7.6	801	8	ADN26185 Bacterial
26	107	7.6	815	7	ABO78505 Pseudomon
27	106.5	7.6	860	9	ADY26642 Streptomy
28	106.5	7.6	3546	4	AAH82212 Polypeptid
29	106.5	7.6	9510	6	AAE36119 Streptomy
30	106	7.5	1641	4	ABG99879 S. cinna
31	106	7.5	3753	4	ABG99878 S. cinna
32	105.5	7.5	152	6	ABU18053 Protein e
33	105.5	7.5	426	7	ABO63304 Klebsiell
34	105.5	7.5	1976	9	ADY80057 Amino aci
35	105.5	7.5	6291	7	AD123927 Streptomy
36	104	7.4	154	5	ABH48055 Listeria
37	104	7.4	4032	8	ADV99897 Nanchangm
38	103	7.3	3025	4	ABG99868 S. cinna
39	103	7.3	5435	4	AAE10145 Streptomy
40	103	7.3	7429	9	ADY80055 Amino aci
41	102.5	7.3	1657	8	ADQ91704 Polypeptid
42	102.5	7.3	1657	8	ABE86996 Streptomy
43	102.5	7.3	1657	8	ABE86793 Streptomy
44	102.5	7.3	10625	9	ADZ51727 PKS FecC
45	102	7.2	455	2	AAE35205 Lipase #2

ALIGNMENTS

RESULT 1  
AD137300 standard; protein; 279 AA.  
AD137300;  
22-APR-2004 (first entry)  
M. tuberculosis low oxygen induced antigen Rv2028c SEQ ID NO:21.  
mycobacterial infection; vaccine; tuberculosis;  
Mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;  
low oxygen induced antigen.  
Mycobacterium tuberculosis.  
WQ2004006952-A2.  
22-JAN-2004.  
08-JUL-2003; 2003WO-DK00477.  
13-JUL-2002; 2002DX-00001098.  
(STAT-) STATENS SERUM INSTR.  
Andersen P, Rosenkrands I, Stryhn A,  
WPI; 2004-122778/12.  
N-PSDB; AD137345.  
Use of one or more polypeptides or their fragments, which are expressed during the latent stage of the mycobacterial infection, and/or nucleic acids encoding the polypeptides, for a therapeutic vaccine against tuberculosis.  
Claim 3; SEQ ID NO 21; 76pp; English.  
The present invention describes polypeptides or their fragments, which are expressed during the latent stage of a mycobacterial infection, and/or nucleic acids encoding the polypeptides, which are useful for creating a therapeutic vaccine against tuberculosis. Also described: (1) a therapeutic vaccine against tuberculosis comprising one or more polypeptides; (2) a method for treating an animal, including a human being, with tuberculosis caused by virulent Mycobacteris, e.g. by Mycobacterium tuberculosis, M. africanum or M. bovis; (3) a method for immunising an animal, including a human being, against tuberculosis

CC caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis  
 CC caused by virulent mycobacteria in an animal, including a human being;  
 CC (5) a method for diagnosing previous or ongoing infection with a virulent  
 CC mycobacterium; and (6) a method of diagnosing Mycobacterium tuberculosis  
 CC infection in a subject. The polypeptides have antibacterial activities,  
 CC and can be used in vaccines and in gene therapy. The polypeptides are  
 CC useful for the manufacture of a therapeutic vaccine for treating an  
 CC individual who is infected by a virulent mycobacterium, e.g. M.  
 CC tuberculosis, and who is not vaccinated with BCG against tuberculosis.  
 CC The present sequence represents a low oxygen induced antigen, which is  
 CC used in the exemplification of the present invention.

XX Sequence 279 AA:

Query Match 100.0%; Score 1410; DB 8; Length 279;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-138;  
 Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNQSHKRPSTIVGIDGSKPAVQALMAVDEASADIPRLIYAEPPDPGTAHGAARK 60  
 DB 1 MNQSHKRPSTIVGIDGSKPAVQALMAVDEASADIPRLIYAEPPDPGTAHGAARK 60  
 QY 61 LAAENAVRYAFVAEADRPVKVEITQERPTSLIRASAAALVCVGAIGVHFRPE 120  
 DB 61 LAAENAVRYAFVAEADRPVKVEITQERPTSLIRASAAALVCVGAIGVHFRPE 120  
 QY 121 RVGSTAAALALSAQCVAIVRPHRVPPIGRDAWIVVEADGSSDIGVLLGAVMAEARLDS 180  
 DB 121 RVGSTAAALALSAQCVAIVRPHRVPPIGRDAWIVVEADGSSDIGVLLGAVMAEARLDS 180  
 QY 181 PVRVVTGROSGVGTGDDVRSALDRWLARMQRPYDVAVOSAAVHGELLDVLAIGRSVH 240  
 DB 181 PVRVVTGROSGVGTGDDVRSALDRWLARMQRPYDVAVOSAAVHGELLDVLAIGRSVH 240  
 QY 241 MVTLSASDQEHVEQLVGAHPNAVLQEAQCTLLVVGQOYL 279  
 DB 241 MVTLSASDQEHVEQLVGAHPNAVLQEAQCTLLVVGQOYL 279

RESULT 2

AD137321 ID AD137321 standard; protein; 268 AA.

AC AD137321;

DT 22-APR-2004 (first entry)

DE M. tuberculosis low oxygen induced antigen Rv1334c SEQ ID NO:42.

KM mycobacterial infection; vaccine; tuberculosis;

KM Mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;

OS low oxygen induced antigen.

XX Mycobacterium tuberculosis.

PN WO2004006952-A2.

PD 22-JAN-2004.

PF 08-JUL-2003; 2003WO-DK000477.

PR 13-JUL-2002; 2002DK-00001098.

PA (STAT-) STATENS SERUM INST.

PI Andersen P, Rosenkrands I, Stryhn A;

XX WPI; 2004-122778/12.

DR N-PSDB; AD137366.

PT Use of one or more polypeptides or their fragments, which are expressed  
 PT during the latent stage of the mycobacterial infection, and/or nucleic  
 PT acids encoding the polypeptides, for a therapeutic vaccine against

PT tuberculosis.

PS Claim 3; SEQ ID NO 42; 76pp; English.

XX The present invention describes polypeptides or their fragments, which  
 XX are expressed during the latent stage of a mycobacterial infection,  
 CC and/or nucleic acids encoding the polypeptides, which are useful for  
 CC creating a therapeutic vaccine against tuberculosis. Also described: (1)  
 CC a therapeutic vaccine against tuberculosis comprising one or more  
 CC polypeptides; (2) a method for treating an animal, including a human  
 CC being, with tuberculosis caused by virulent mycobacteria, e.g. by  
 CC Mycobacterium tuberculosis, M. africanum or M. bovis; (3) a method for  
 CC immunising an animal, including a human being, against tuberculosis  
 CC caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis  
 CC caused by virulent mycobacteria in an animal, including a human being;  
 CC (5) a method for diagnosing previous or ongoing infection with a virulent  
 CC mycobacterium; and (6) a method of diagnosing Mycobacterium tuberculosis  
 CC infection in a subject. The polypeptides have antibacterial activities,  
 CC and can be used in vaccines and in gene therapy. The polypeptides are  
 CC useful for the manufacture of a therapeutic vaccine for treating an  
 CC individual who is infected by a virulent mycobacterium, e.g. M.  
 CC tuberculosis, and who is not vaccinated with BCG against tuberculosis.  
 CC The present sequence represents a low oxygen induced antigen, which is  
 CC used in the exemplification of the present invention.

XX Sequence 268 AA:

Query Match 30.6%; Score 431.5; DB 8; Length 268;  
 Best Local Similarity 40.4%; Pred. No. 4.3e-36;  
 Matches 111; Conservative 34; Mismatches 105; Indels 25; Gaps 8;

QY 1 MNQSHKRPSTIVGIDGSKPAVQALMAVDEASADIPRLIYAEPPDPGTAHGAARK 60

DB 1 MSDPRPARAVVVGIDGSKPAVQALMAVDEAVRDIPLRLVYIDISQLSAAGEGGO-- 58

QY 61 LAAENAVRYAFVAEADRPVKVEITQERPTSLIRASAAALVCVGAIGVHFRPE 120

DB 59 -SARAAALHDASKKVAETGVPKIEVLCGRPLTKMQSSRAALCGSGVLDIVRGR 117

QY 121 RVGSTAAALALSAQCVAIVRPHRVPPIGRDAWIVVEADGSSDIGVLLGAVMAEARL 178

DB 118 R-GSVAATLALGSLCVAIVHPSRAPATTSSQSAVAVAVDN---GVTLRHAFFEARLR 172

QY 179 DSPVRVVTGROSGVGTGDDV-----RASLDRLWLARMQRPYDVAVOSAAVHGELLDV 231

DB 173 GVPLRAVAVH---AETPPDVEGSRILAHVHLSRRLAHHTRLTYREVRVRAIAGSACRH 229

QY 232 LAGLGRSVHNVVLSADQEHVEQLVGA--PGNAVLT 264

DB 230 LAANAKPGQLFV--ADSHSAHELCAVOPGCAVL 261

RESULT 3

AD137306 ID AD137306 standard; protein; 272 AA.

AC AD137306;

DT 22-APR-2004 (first entry)

DE M. tuberculosis low oxygen induced antigen Rv2624c SEQ ID NO:27.

KM mycobacterial infection; vaccine; tuberculosis;

KM Mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;

OS low oxygen induced antigen.

XX Mycobacterium tuberculosis.

PN WO2004006952-A2.

PD 22-JAN-2004.

PF 08-JUL-2003; 2003WO-DK000477.



XX 13-JUL-2002; 2002DK-00001098.  
 PR (STAT-) STATENS SERUM INST.  
 PA Andersen P, Rosenkrands I, Stryhn A;  
 PI WPI; 2004-122778/12.  
 DR N-PSDB; ADI37351.  
 XX  
 PT Use of one or more polypeptides or their fragments, which are expressed  
 PT during the latent stage of the mycobacterial infection, and/or nucleic  
 PT acids encoding the polypeptides, for a therapeutic vaccine against  
 PT tuberculosis.  
 PS  
 PS Claim 3; SEQ ID NO 27; 76pp; English.  
 CC The present invention describes polypeptides or their fragments, which  
 CC are expressed during the latent stage of a mycobacterial infection,  
 CC and/or nucleic acids encoding the polypeptides, which are useful for  
 CC creating a therapeutic vaccine against tuberculosis. Also described: (1)  
 CC a therapeutic vaccine against tuberculosis comprising one or more  
 CC polypeptides; (2) a method for treating an animal, including a human  
 CC being, with tuberculosis caused by virulent mycobacteria, e.g. by  
 CC Mycobacterium tuberculosis, M. africanum or M. bovis; (3) a method for  
 CC immunising an animal, including a human being, against tuberculosis  
 CC caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis  
 CC caused by virulent mycobacteria in an animal, including a human being;  
 CC (5) a method for diagnosing previous or ongoing infection with a virulent  
 CC mycobacterium; and (6) a method of diagnosing Mycobacterium tuberculosis  
 CC infection in a subject. The polypeptides have antibacterial activities,  
 CC and can be used in vaccines and in gene therapy. The polypeptides are  
 CC useful for the manufacture of a therapeutic vaccine for treating an  
 CC individual who is infected by a virulent mycobacterium, e.g. M.  
 CC tuberculosis, and who is not vaccinated with BCG against tuberculosis.  
 CC The present sequence represents a low oxygen induced antigen, which is  
 CC used in the exemplification of the present invention.  
 CC  
 XX Sequence 272 AA:  
 SO  
 Query Match 29.5%; Score 416; DB 8; Length 272;  
 Best Local Similarity 36.8%; Pred. No. 1.8e-34;  
 Matches 98; Conservative 46; Mismatches 116; Indels 6; Gaps 2;  
 QY 9 SIIVGIDGSRPAVOALMAVDEASRDIPRLVLAIPDDPGVAHGAARKLAENAV 68  
 DB 11 TIVGIDGSHAAITPAALMGVDEATSRAPLVIVIRTPSPDYD--RDLHAERSL 67  
 QY 69 RYAFATAEAADRPKVEVEITQERPTSLIRASAAALVCGAIGVHFRPERVSTPAA 128  
 DB 68 RMAQSAVEAAGKLVKTIETDIPRGPAGPVIVASRDAMICGSGICIRVASSILGSTATE 127  
 QY 129 LAISAQCPVAIVRPHRVPIGRDAAMIVVEADGSSDIGVLLGAVMAEARLRDSPVRYTCR 188  
 DB 128 LAEKARCPVAVMRSKVDQAPASDINWIMTADPNDAVAVLEYARERAKLQAAPITALGCR 187  
 QY 189 QSGVGIDGDDVRASLSDRWLARWQRPYRVQSAAVGELLDYAGIGRSYHMYVLSASD 246  
 DB 188 PEBIRRETPD--GEFERRVODMHRHPDVRYPIITHTGIRFLADHDERVQLAVIGGE 244  
 QY 249 QEHVEQLVGAAGNAVLOEAGCTLLVV 274  
 DB 245 AGQALRVGSPGHVFRHAECSVLVV 270  
 RESULT 4  
 ADI37298  
 ID ADI37298 standard; protein; 295 AA.  
 AC ADI37298;  
 XX  
 XX 22-APR-2004 (first entry)  
 XX

DE M. tuberculosis low oxygen induced antigen Rv2005c SEQ ID NO:19.  
 XX  
 XX mycobacterial infection; vaccine; tuberculosis;  
 KM Mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;  
 KM low oxygen induced antigen.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN W02004006952-A2.  
 XX  
 PD 22-JAN-2004.  
 XX  
 XX 08-JUL-2003; 2003WO-DK000477.  
 XX  
 XX 13-JUL-2002; 2002DK-00001098.  
 PR  
 PA (STAT-) STATENS SERUM INST.  
 PI Andersen P, Rosenkrands I, Stryhn A;  
 DR WPI; 2004-122778/12.  
 DR N-PSDB; ADI37343.  
 XX  
 XX Use of one or more polypeptides or their fragments, which are expressed  
 XX during the latent stage of the mycobacterial infection, and/or nucleic  
 XX acids encoding the polypeptides, for a therapeutic vaccine against  
 XX tuberculosis.  
 PS  
 PS Claim 3; SEQ ID NO 19; 76pp; English.  
 CC The present invention describes polypeptides or their fragments, which  
 CC are expressed during the latent stage of a mycobacterial infection,  
 CC and/or nucleic acids encoding the polypeptides, which are useful for  
 CC creating a therapeutic vaccine against tuberculosis. Also described: (1)  
 CC a therapeutic vaccine against tuberculosis comprising one or more  
 CC polypeptides; (2) a method for treating an animal, including a human  
 CC being, with tuberculosis caused by virulent mycobacteria, e.g. by  
 CC Mycobacterium tuberculosis, M. africanum or M. bovis; (3) a method for  
 CC immunising an animal, including a human being, against tuberculosis  
 CC caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis  
 CC caused by virulent mycobacteria in an animal, including a human being;  
 CC (5) a method for diagnosing previous or ongoing infection with a virulent  
 CC mycobacterium; and (6) a method of diagnosing Mycobacterium tuberculosis  
 CC infection in a subject. The polypeptides have antibacterial activities,  
 CC and can be used in vaccines and in gene therapy. The polypeptides are  
 CC useful for the manufacture of a therapeutic vaccine for treating an  
 CC individual who is infected by a virulent mycobacterium, e.g. M.  
 CC tuberculosis, and who is not vaccinated with BCG against tuberculosis.  
 CC The present sequence represents a low oxygen induced antigen, which is  
 CC used in the exemplification of the present invention.  
 CC  
 XX Sequence 295 AA:  
 SO  
 Query Match 18.2%; Score 257; DB 8; Length 295;  
 Best Local Similarity 29.4%; Pred. No. 7.2e-18;  
 Matches 91; Conservative 43; Mismatches 126; Indels 50; Gaps 9;  
 QY 1 MNSHSPSIVGIDGSRPAVOALMAVDEASRDIPRLVLAIPD----- 47  
 DB 1 MSKPRQHGIVGVVDLSLSDAAACGATDAARNIPLVVHVNVNADVATWPMYPTW 60  
 QY 48 -----DPGYAAHGAARKLAENAVRYAFATAEAADRPKVEVEITQERPTSLIRASA 102  
 DB 61 GVMQEDBG-----RQIYA--NAKRLAKEAV-GADRKLSVSVSELVFTSPVPTWELSN 109  
 QY 103 AALVCGAIGVHFRPERVSTPAAALALSAQCPVAIVR-----PHEVP--IGRDAA 152  
 DB 110 EAEWVVLGSSGRGALARGLVSSSVLVRACGPVAVIHSDDAVIPPQIAPVLVIGDGS 169  
 QY 153 WI-----VVEADSSSDIGVLLGAVMAEARLRDSPVRYVTCRQSGVDPTGDVVAASLDRL 207  
 DB 170 PVSELATPAVAFDEASRGVELLAVHAW-----SDVEVELPGLDFAVQOGEALSLAERL 224

QY 208 ARWQPRYPDVRSAAVHGBLLDYLAGRSVHVVLSASDQEHVF-QLYGAPGNAYLQE 266  
 DB 225 AGMQERYPDVRSRVVVCDBRPARKLVOKSASAOQLVVGSHRGGLTGMLIGSVSNAYLHA 284  
 QY 267 AGCTLLVVGQ 276  
 DB 285 ARVPVIVARQ 294  
 RESULT 5  
 ID AAY87878 standard; protein; 297 AA.  
 XX AAY87878;  
 AC AAY87878;  
 DT 06-OCT-2000 (first entry)  
 XX M. tuberculosis antigen TB32 protein.  
 DE M. tuberculosis antigen TB32 protein.  
 XX Tuberculosis; TB; antigen; vaccine; diagnosis; somatic; tuberculostatic;  
 KM infection; interferon-gamma; IFN-gamma; protective immunity; therapy;  
 KM delayed type hypersensitivity response; TB32.  
 XX Mycobacterium tuberculosis.  
 OS  
 PN WO200021983-A2.  
 PD 20-APR-2000.  
 XX 08-OCT-1999; 99WO-DK000538.  
 PF 08-OCT-1998; 98DK-00001281.  
 PR 21-JAN-1999; 99US-0116673P.  
 XX (STAT-) STATENS SERUM INST.  
 PA  
 XX Andersen P, Melding K, Hansen CV, Florio W, Okkels LMM;  
 PI Skjot RLV, Rosenkrands I;  
 DR N-PSDB; AAA39572.  
 XX WPI; 2000-317931/27.  
 PT Novel polypeptide of somatic protein extract useful as vaccine against  
 PT virulent Mycobacterium infection. Isolated from cell wall, cell membrane  
 PT and cytosol.  
 PS Claim 1; Page 106-107; 126pp; English.  
 CC This invention describes a novel polypeptide (PP) of somatic proteins  
 CC extract (I) which have tuberculostatic activity. (I) or their subsequence  
 CC has at least one of the following properties: (a) the PP induces an in  
 CC vitro recall response, or an in vitro response, during primary infection  
 CC with virulent Mycobacterium, determined by a release of interferon (IFN)-  
 CC gamma. (b) PP induces a protective immunity, determined by vaccinating an  
 CC animal with PP and an adjuvant, three times at two weeks intervals, (c)  
 CC PP induces an in vitro response, or in vitro recall response, determined  
 CC by release of IFN-gamma of at least 1000 pg/ml or 500 pg/ml,  
 CC respectively, from peripheral blood mononuclear cells (PBMC) withdrawn  
 CC from TB patients, or PPD positive individuals, 6 months after diagnosis,  
 CC (d) PP induces a specific antibody response in a TB patient, as  
 CC determined by enzyme linked immunosorbent assay (ELISA) technique or a  
 CC western blot. (e) PP induces a positive delayed type hypersensitivity  
 CC (DTH) response, determined by intradermal injection. (I) and (II) are  
 CC useful in preparing a prophylactic or therapeutic medicine as a vaccine  
 CC for induction of a protective or generation of an immune response in a  
 CC mammal against infection with a virulent Mycobacterium. (I) and (II) are  
 CC also useful as diagnostic reagent for the diagnosis of a virulent  
 CC Mycobacterium infection. The vaccine of the invention induces efficient  
 CC immunological memory, providing long term protection against TB. This  
 CC sequence represents a Microbacterium tuberculosis TB32 antigen described  
 CC in the invention  
 CC Sequence 297 AA;  
 SQ

Query Match 15.2%; Score 214; DB 3; Length 297;  
 Best Local Similarity 27.2%; Pred. No. 2.2e-13;  
 Matches 84; Conservative 37; Mismatches 134; Indels 54; Gaps 10;  
 QY 1 MNQSHKPPISIVGIDGSKPAVQAALMAVDEASRDILRLVLAEPD-----DPQY 51  
 DB 1 MSSGNSISLIGIVGIDSPAAQVAVRWAAADAEIRKIPLTGVHNVSPVATWLEVPILPGV 60  
 QY 52 AA---HGAARKLAENAVRYAFPAVEAAD---RPVVEVEITQERPVTSILIRASAA 104  
 DB 61 LRWQDGHG-----RHLDALKVVEQASLRACRPVHSEIVPAAPVTLVDMSKXA 111  
 QY 105 ALVCVGAIGVHHFRPVRVSTAAALASACCPVALV-----RPHRVPIGRDAAMIV 156  
 DB 112 VLMVVGCLSGRWPRGLTGSVSGILRHACFPVITIHEDSVMPHFQAP-----VLV 164  
 QY 157 EADGSSDIGVLGAVVAEARKLDRSPRVVYTCRQSGGDF-----GDVRS-----LDR 205  
 DB 165 GVDGSSASELATRIAFDEASRRN--VDLVALLHMSDVDSWMPGIDWPATQSMACQVLA 222  
 QY 206 WLRWQPRYPDVRSAAVHGBLLDYLAGRSVHVVLSASDQ-EHVEQLVGAPGNAYL 264  
 DB 223 RLAWQERYPNVAITRVVTRDQPARQLVGRSEBAQLVVGSRGRGYSACMLVGSGETVA 282  
 QY 265 QEAGCTLLV 273  
 DB 283 QLARTPVIV 291  
 RESULT 6  
 ABG30907  
 ID ABG30907 standard; protein; 297 AA.  
 XX ABG30907;  
 AC  
 XX 21-OCT-2002 (first entry)  
 DT  
 XX Mycobacterium tuberculosis dormancy-induced protein Rv2623.  
 DE  
 XX Anti-mycobacterial; Mycobacterium tuberculosis; strain H37Rv;  
 KM non-oxygen limiting; hypoxic stationary; hypoxic growth phase;  
 KM tuberculosis; dormant mycobacterial infection; dormancy-induced; Rv2623;  
 KM tuberculostatic; vaccine.  
 XX Mycobacterium tuberculosis.  
 OS  
 PN WO200248391-A2.  
 PD 20-JUN-2002.  
 XX 11-DEC-2001; 2001WO-EP014551.  
 PR 13-DEC-2000; 2000GB-00030368.  
 XX (MOLE-) INST MOLECULAR & CELL BIOLOGY.  
 PA  
 XX Dick T, Calvin BKK;  
 PI WPI; 2002-583530/62.  
 DR N-PSDB; ABK9546.  
 XX Identifying an anti-mycobacterial agent that modulates  
 XX PT activity/expression of a protein expressed by Mycobacterium, involves  
 XX PT monitoring the effect of an agent on the activity/expression of the  
 XX PT protein or polynucleotide/vector encoding it.  
 PS Example 2; Page 44-45; 56pp; English.  
 CC The present invention relates to a new method for the identification of  
 CC anti-mycobacterial agent that modulates activity and/or expression of  
 CC protein expressed by Mycobacterium in non-oxygen limiting or hypoxic  
 CC stationary, hypoxic growth phase. The method involves contacting a test



XX	Sequence	301 AA:
XX	Query Match	14.7%; Score 207; DB 4; Length 301;
XX	Best Local Similarity	26.1%; Pred. No. 1.2e-12;
XX	Matches	77; Conservative 46; Mismatches 134; Indels 38; Gaps 10
QY	10 IVVGIDGSKAVQALMALNDAEASRIPIRL-----LYA---IFPDPPGIAAHGA 57	
DB	7 VVAVVDSDSDASKQAVRMAANTANKRIPIRLRLSYTTPQFILAEGMVPQELPFDLQAEA 66	
QY	58 ARKLAAEKNVRAFTAVEAADRPVVEVEITQERPVTLIRASAAALVCVGAIGVHHF 117	
DB	67 LEKINERAD-----IAHEVAPE-IKIGTTIAEGSPIDMLEMSDAMTWMSRGIGGL 119	
QY	118 RPERVGSSTAALALSAOCPPVAIVRPRVPIGRDAAM--IVEADGSSDIGVILGAVAAE 175	
DB	120 SGWVGSGSVAIVSHAKCPVVVVREDSDA-VNEDSKXGPPVVGVDGSEVSGQATEYAFAEA 178	
QY	176 RLR-----DSRPVRYTCRQSGVGDGDDV-RASLDRMLARKMP---RIPDVRVQ 222	
DB	179 EARGAEIVAVHTWMDQVQASLAGLAAQQQDVEVERQCTDMLIERLAPLVEKTPSVTVK 238	
QY	221 SAAVHGEILLYTLGLGRSVYMWVTLASDQEHVE-OLVGAFGNAVLOEAGCTLLVAV 274	
DB	239 KIITRDRPVRLAEASBNQILVSGHGKGRKMLGSTSRALLQASAPCEPMVV 293	
RESULT 8		
AAU71872	AAU71872 standard; protein; 301 AA.	
ID	AAU71872	
XX	AAU71872;	
AC	26-FEB-2002 (first entry)	
DT		
XX		
DE	C. glutamicum metabolic pathway protein encoded by gene #7.	
XX		
KW	Metabolic pathway protein; MP; lysine biosynthesis pathway;	
KW	methionine biosynthesis pathway; large-scale production of fine chemical;	
XX	Corynebacterium diphtheriae; diphtheria.	
OS	Corynebacterium glutamicum.	
XX		
PN	WO20016573-A2.	
PD		
XX	13-SEP-2001.	
XX		
PF	22-DEC-2000; 2000MCO-IB002035.	
XX		
PR	09-MAR-2000; 2000US-0187970P.	
PR	23-JUN-2000; 2000US-00606740.	
XX		
PA	(BADI ) BASF AG.	
XX		
P1	Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;	
P1	Kim J, Lee H, Hwang B;	
DR	WPI; 2001-582269/65.	
DR	N-PSDB; AAS96082.	
XX		
PT	Nucleic acids encoding metabolic pathway proteins from Corynebacterium	
PT	glutamicum, useful for producing methionine and lysine in Corynebacterium	
PT	and Brevibacterium.	
XX		
PS	Disclosure; Page 169-170; 316pp; English.	
XX		
CC	The present invention relates to the isolation of novel Corynebacterium	
CC	glutamicum genes (AAS96073-AAS96132) encoding metabolic pathway (MP)	
CC	proteins. The metabolic pathway proteins of the invention include enzymes	
CC	involved in the lysine and methionine biosynthetic pathways. The	
CC	polynucleotide sequences of the invention can be used for the large-scale	
CC	production and/or modulation of expression of fine chemicals such as	

[illegible]

CC proteins. The metabolic pathway proteins of the invention include enzymes  
 CC involved in the lysine and methionine biosynthetic pathways. The  
 CC polynucleotide sequences of the invention can be used for the large-scale  
 CC production and/or modulation of expression of fine chemicals such as  
 CC lysine and methionine. The sequences of the invention may be used to  
 CC identify C. glutamicum and related organisms e.g. C. diphtheriae in a  
 CC subject to detect diphtheria. AAU71863-AAU71922 represent the novel C.  
 CC glutamicum metabolic pathway proteins of the invention  
 XX

SO Sequence 301 AA;

Query Match 14.7%; Score 207; DB 4; Length 301;  
 Best Local Similarity 26.1%; Pred. No. 1.2e-12;

Matches 77; Conservative 46; Mismatches 134; Indels 38; Gaps 10;

QY 10 IVVGIDGSKPAVQALMAVDEAASRDIPRL-----LYA---IEPDDPGYAAGAA 57  
 DB 7 VVAVVDSGSDASKQAVRMAANTANKRGIPRLASSTYMPQFLYAGMTVPQELFPDLOAEA 66  
 QY 58 ARKLAAMENAVRYAFTAVEAADRPVKEVEITQERPTSLIRASAAALVCVGAIGVHHF 117  
 DB 67 LEKINEARD-----IAHEVAPE-IKIGHTIAEGSPIDMLLEMSPDATMIWMSRGIGGL 119  
 QY 118 RPEKVGSTAAALALSAQCPVALVRPHRVPFGRDAA--IYVEADGSSDITGVLGAWMAEA 175  
 DB 120 SGWVGSGVSGAVVSHAKCPVVVVEDSA-VNEDSKYGPVVVGVDGSEVSGQATEYAFPAEA 178  
 QY 176 RLR-----DSPRVVTCRQSGVGTGDDV--RASLDRLMARQOP---RYPDVRVQ 220  
 DB 179 EARGAEIVAVHTMDMVQASLAGLAAAOQOMDEVERQCTDMLIERLAPLVEKIPSTYVK 238  
 QY 221 SAAVHGELIDYLAGRSVHMVVLASDQEHVE-QLVAGPAGNAVLOEAGCTLLVV 274  
 DB 239 KIIRDRPVRALAEASENAQLLVVSGHGRGFGKMLGSTRALLQSAPCPMVV 293

RESULT 10  
 AAG92991  
 ID AAG92991 standard; protein; 301 AA.

XX AAG92991;

XX 26-SEP-2001 (first entry)

XX C glutamicum protein fragment SEQ ID NO: 6745.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

XX organic acid synthesis.

XX Corynebacterium glutamicum.

XX EPI108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-00127688.

XX 16-DEC-1999; 99JP-00377484.

XX 07-APR-2000; 2000JP-00159162.

XX 03-AUG-2000; 2000JP-00280988.

XX (KYOWA ) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX MPI; 2001-376931/40.

XX N-PSDB; AAH68210.

PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analyzing  
 PT expression profile or pattern of a gene and identifying homologous gene.

PS Claim 17; SEQ ID NO 6745; 246bp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and analyzing  
 CC the expression profile or expression pattern of a gene derived from  
 CC Coryneform bacterium, and identifying a homologue of a gene derived from  
 CC coryneform bacterium. Coryneform bacteria are useful for producing amino  
 CC acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described in the  
 CC exemplification of the invention. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in  
 CC electronic format directly from the European Patent Office  
 XX

SO Sequence 301 AA;

Query Match 14.7%; Score 207; DB 4; Length 301;  
 Best Local Similarity 26.1%; Pred. No. 1.2e-12;

Matches 77; Conservative 46; Mismatches 134; Indels 38; Gaps 10;

QY 10 IVVGIDGSKPAVQALMAVDEAASRDIPRL-----LYA---IEPDDPGYAAGAA 57  
 DB 7 VVAVVDSGSDASKQAVRMAANTANKRGIPRLASSTYMPQFLYAGMTVPQELFPDLOAEA 66  
 QY 58 ARKLAAMENAVRYAFTAVEAADRPVKEVEITQERPTSLIRASAAALVCVGAIGVHHF 117  
 DB 67 LEKINEARD-----IAHEVAPE-IKIGHTIAEGSPIDMLLEMSPDATMIWMSRGIGGL 119  
 QY 118 RPEKVGSTAAALALSAQCPVALVRPHRVPFGRDAA--IYVEADGSSDITGVLGAWMAEA 175  
 DB 120 SGWVGSGVSGAVVSHAKCPVVVVEDSA-VNEDSKYGPVVVGVDGSEVSGQATEYAFPAEA 178  
 QY 176 RLR-----DSPRVVTCRQSGVGTGDDV--RASLDRLMARQOP---RYPDVRVQ 220  
 DB 179 EARGAEIVAVHTMDMVQASLAGLAAAOQOMDEVERQCTDMLIERLAPLVEKIPSTYVK 238  
 QY 221 SAAVHGELIDYLAGRSVHMVVLASDQEHVE-QLVAGPAGNAVLOEAGCTLLVV 274  
 DB 239 KIIRDRPVRALAEASENAQLLVVSGHGRGFGKMLGSTRALLQSAPCPMVV 293

RESULT 11  
 ABG80324  
 ID ABG80324 standard; protein; 301 AA.

XX ABG80324;

XX 15-NOV-2002 (first entry)

XX C. glutamicum metabolic pathway (MP) protein #4.

XX Metabolic pathway protein; MP; cell metabolism; amino acid; vitamin;

XX cofactor; nucleotide; nucleoside; trehalose; fine chemical production;

XX organic acid; non-proteinogenic amino acid; purine base; carbohydrate;

XX pyrimidine base; lipid; unsaturated fatty acid; diol; polyketide;

XX aromatic compound; food industry; animal feed; cosmetic industry;

XX pharmaceutical industry; enzyme.

XX Corynebacterium glutamicum ATCC 13032.

XX MO200251231-A1.

XX 04-JUL-2002.

XX 22-DEC-2000; 2000WO-EP013143.

XX 22-DEC-2000; 2000WO-EP013143.

XX (BADT ) BASF AG.

XX Pompejus M, Kroegeer B, Zelder O, Schroeder H;

DR WPI; 2002-643289/69.  
 DR N-PSDB; ABB5345.  
 XX  
 PT New metabolic pathway genes of *Corynebacterium glutamicum* for producing  
 PT fine chemicals, e.g. lipids, (unsaturated fatty acids, vitamins,  
 PT cofactors or enzymes used in food, feed, cosmetics or pharmaceutical  
 PT industries.  
 XX  
 XX Claim 18; Page 107-108; 176pp; English.  
 XX  
 CC The present invention relates to the isolation of *Corynebacterium*  
 CC glutamicum metabolic pathway (MP) proteins, and the polynucleotide  
 CC sequences encoding them. The MP proteins are enzymes involved in the  
 CC metabolism of molecules important for the normal functioning of cells  
 CC (e.g. amino acids, vitamins, cofactors, nucleotides and nucleosides, or  
 CC trehalose). The polynucleotide sequences encoding the MP proteins are  
 CC useful for producing fine chemicals, particularly organic acids, non-  
 CC proteinogenic amino acids, purine and pyrimidine bases, nucleosides,  
 CC nucleotides, lipids, (unsaturated fatty acids, diols, carbohydrates,  
 CC aromatic compounds, vitamins, cofactors, polycyclics and enzymes. The  
 CC fine chemicals are useful in the food, animal feed, cosmetic or  
 CC pharmaceutical industries. ABB5345-ABB5343 represent the C. glutamicum  
 CC MP proteins of the invention  
 CC  
 SQ Sequence 301 AA;  
 XX  
 Query Match 14.7%; Score 207; DB 5; Length 301;  
 Best Local Similarity 26.1%; Pred. No. 1.2e-12;  
 Matches 77; Conservative 46; Mismatches 134; Indels 38; Gaps 10;  
 QY 10 IVVGIGDSKRAVQALMAVDEASRDIPRL-----LYA---IEPDDPGYAHGAA 57  
 DB 7 VVAVVDSDASKQAVRWAAANTANKRGIPRLASSTYMPQFLYAGVWPQELFDLQAEA 66  
 QY 58 ARKLAAMENAVRFTAVEAADRPVKEVEITQERPTSLIRASAAALVCVGAIGVHHF 117  
 DB 67 LEKINEARD-----IAHEVAPE-IKGTHTAAGSPIDMLIEMSPDXTMIVGSRIGL 119  
 QY 118 RPERVGSSTAALALSAQCPSVAIVRPHRVPICGRDAW--IVEADGSSDIGVLGAVMAEA 175  
 DB 120 SGWVMSVSGAVVSHAKCPVVVVEDSA-VNEDSKGPPVVGVDGSEVSQATEYAFAEA 178  
 QY 176 RLR-----DSPRVVTCRQSGVGTGDDV-RASLDRLAKMP---RYPDVRYQ 220  
 DB 179 EARGAEIVAVHTWMDQVQASLAGIAAAQOQWDEVERQOTDMLIERLAPLVEKTPSYTVK 238  
 QY 221 SAAVHGELDYLAGRSVHMVVLASDQEHVE-QLVGAPGNAYLQAGCTLLV 274  
 DB 239 KIITDRPVPALAEASNAQLLVGSHRGSGFKMLIGSTSRALLQSAFCPMVV 293  
 RESULT 12  
 ABB5399 standard; protein; 301 AA.  
 XX  
 AC ABB5399;  
 XX  
 DT 22-SEP-2005 (first entry)  
 XX  
 DE C glutamicum metabolic pathway regulatory (MR) protein Segid314.  
 XX  
 KM metabolism; microorganism detection; microorganism identification;  
 KM genome; evolution; protein production.  
 XX  
 OS *Corynebacterium glutamicum*.  
 XX  
 PN US2005153402-A1.  
 XX  
 PD 14-JUL-2005.  
 XX  
 PF 06-DEC-2004; 2004US-00006098.  
 XX  
 PR 25-JUN-1999; 99US-0141031P.

PR 01-JUL-1999; 99DE-01030476.  
 PR 01-JUL-1999; 99US-0142690P.  
 PR 08-JUL-1999; 99DE-01031419.  
 PR 08-JUL-1999; 99DE-01031420.  
 PR 09-JUL-1999; 99DE-01032122.  
 PR 09-JUL-1999; 99DE-01032128.  
 PR 09-JUL-1999; 99DE-01032134.  
 PR 09-JUL-1999; 99DE-01032206.  
 PR 09-JUL-1999; 99DE-01032207.  
 PR 14-JUL-1999; 99DE-01033003.  
 PR 27-AUG-1999; 99US-0151251P.  
 PR 31-AUG-1999; 99DE-0141390.  
 PR 03-SEP-1999; 99DE-01042088.  
 PR 03-SEP-1999; 99DE-01042124.  
 PR 23-JUN-2000; 2000US-00602874.  
 PA (BADI ) BASF AG.  
 XX  
 PI Pompejus M, Kroger B, Schroder H, Zelder O, Haberhauer G;  
 XX  
 DR WPI; 2005-496831/50.  
 DR N-PSDB; ABB15398.  
 XX  
 PT An isolated metabolic pathway regulatory polypeptide from *Corynebacterium*  
 PT glutamicum, its portion, or its variant, useful for diagnosing presence  
 PT or activity of C. diphteriae in subject.  
 XX  
 XX Claim 20; SEQ ID NO 314; 65pp; English.  
 PS  
 CC This invention relates to novel isolated metabolic pathway regulatory  
 CC polypeptides from *Corynebacterium glutamicum*, designated MR proteins, and  
 CC the DNA sequences which encode them. The invention is useful in  
 CC identification of C glutamicum and related organisms, mapping of genomes  
 CC of organisms related to C glutamicum, identification and localization of  
 CC C glutamicum sequences of interest, evolutionary studies, determination  
 CC of MR protein regions required for function, modulation of MR protein  
 CC activity and modulation of cellular production of desired compounds such  
 CC as a fine chemical. The present sequence is that of a novel isolated  
 CC metabolic pathway regulatory polypeptide of the invention from  
 CC *Corynebacterium glutamicum*. Note: This sequence does not appear in the  
 CC specification but was obtained in electronic format directly from USPTO.  
 CC  
 SQ Sequence 301 AA;  
 XX  
 Query Match 14.7%; Score 207; DB 9; Length 301;  
 Best Local Similarity 26.1%; Pred. No. 1.2e-12;  
 Matches 77; Conservative 46; Mismatches 134; Indels 38; Gaps 10;  
 QY 10 IVVGIGDSKRAVQALMAVDEASRDIPRL-----LYA---IEPDDPGYAHGAA 57  
 DB 7 VVAVVDSDASKQAVRWAAANTANKRGIPRLASSTYMPQFLYAGVWPQELFDLQAEA 66  
 QY 58 ARKLAAMENAVRFTAVEAADRPVKEVEITQERPTSLIRASAAALVCVGAIGVHHF 117  
 DB 67 LEKINEARD-----IAHEVAPE-IKGTHTAAGSPIDMLIEMSPDXTMIVGSRIGL 119  
 QY 118 RPERVGSSTAALALSAQCPSVAIVRPHRVPICGRDAW--IVEADGSSDIGVLGAVMAEA 175  
 DB 120 SGWVMSVSGAVVSHAKCPVVVVEDSA-VNEDSKGPPVVGVDGSEVSQATEYAFAEA 178  
 QY 176 RLR-----DSPRVVTCRQSGVGTGDDV-RASLDRLAKMP---RYPDVRYQ 220  
 DB 179 EARGAEIVAVHTWMDQVQASLAGIAAAQOQWDEVERQOTDMLIERLAPLVEKTPSYTVK 238  
 QY 221 SAAVHGELDYLAGRSVHMVVLASDQEHVE-QLVGAPGNAYLQAGCTLLV 274  
 DB 239 KIITDRPVPALAEASNAQLLVGSHRGSGFKMLIGSTSRALLQSAFCPMVV 293  
 RESULT 13  
 AAU62080 standard; protein; 316 AA.  
 XX

[illegible]

Db	165	TSVAAGKSAAPVVVITSPASTPTDPTGKKCAVVAAGPQSVGAIAVGFGRALADHRCKLL	224
Qy	184	VTCQSGVGTGDDVRASLDRLMARQPRYEDVVGQSAVAHGELLDYLAGGRSVHVV	243
Db	225	AVTVRGNDSEEHDBALKRLNEVVAEP LANKPHHVEVERVLSGEPRVDLVLDSGNVDLLV	284
Qy	244	LSASDQHEVEQLVGAFGNVAVLQEGCTLLV	273
Db	285	IGMKKHPILGWTAGGVSPRAIMAHQSPILA I	314
RESULT 14			
ABMS58599			
ID	ABMS58599	standard; protein; 316 AA.	
XX			
AC	ABMS58599;		
XX			
DT	20-OCT-2003	(first entry)	
XX			
DE	Propionibacterium acnes	predicted ORF-encoded polypeptide #23275.	
XX			
KM	Acne vulgaris; antiseborrheic; dermatological; antibacterial;		
XX	immunostimulant; immune response; vaccine.		
XX	Propionibacterium acnes.		
OS			
XX	WO2003033515-A1.		
PN			
XX			
PD	24-APR-2003.		
XX			
PF	11-OCT-2002; 2002WO-US032727.		
XX			
PR	15-OCT-2001; 2001US-00978825.		
XX			
PA	(CORI-) CORIXA CORP.		
XX			
PI	Mitcham JI, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JI;		
XX	Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;		
PI	Barch B, Vallieve-Dougllass J;		
XX			
DR	WPI, 2003-381789/36.		
XX			
XX	N-PsDB; ACP64552.		
PS			
XX	Example 1; SEQ ID NO 23275; 1481bp; English.		
XX			
CC	The invention relates to an isolated polynucleotide (ACF64435-ACP64733)		
CC	encoding a Propionibacterium acnes protein. The invention also relates to		
CC	polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to		
CC	immunogenic fragments of P. acnes polypeptides. The invention		
CC	additionally encompasses expression vectors and host cells comprising a		
CC	polynucleotide of the invention; antibodies against polypeptides of the		
CC	invention; fusion proteins comprising a polypeptide of the invention; a		
CC	method for stimulating an immune response specific for a P. acnes		
CC	polypeptide and an isolated T cell population comprising T cells prepared		
CC	via this method; a vaccine composition (comprising P. acnes polypeptides,		
CC	polynucleotides, antibodies, fusion proteins, T cell populations, or		
CC	antigen-presenting cells that express the polypeptide); a method and kit		
CC	for detecting or determining the presence or absence of P. acnes in a		
CC	patient; and a method for inhibiting the development of P. acnes in a		
CC	patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion		
CC	proteins, T cell populations or antigen-presenting cells that express the		
CC	polypeptides are useful for diagnosing, preventing or treating acne		
CC	vulgaris, or for stimulating an immune response specific for a P. acnes		
CC	protein. The polynucleotides can also be used as probes or primers for		
CC	nucleic acid hybridization. The vaccine composition is useful for the		
CC	stimulation of an immune response against P. acnes, or for treating acne,		
CC	and the kit is useful for performing a diagnostic assay. The present		
CC	sequence represents a polypeptide predicted to be encoded by an ORF (open		



CC reading frame) contained within the P. acnes polymucleotides of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 316 AA;

Query Match 13.8%; Score 194; DB 6; Length 316;  
Best Local Similarity 25.2%; Pred. No. 2.8e-11;  
Matches 68; Conservative 43; Mismatches 151; Indels 8; Gaps 3;

QY 10 IVVGIDSKPAVQALMAVDEAASRDIPRLVLAIEPDDPGYAHGAARLTAENAVR 69  
DB 47 IIVGVDSSEGLRARAAYAGSARIDDLVLHAV--DDAAVACAMGVVDPYTLQDAQ 104  
QY 70 YA-----FTAVEADRPVKVEVEITQERPRVTLIRASAAALVCVGAIGVHFRPRVGS 124  
DB 105 VVVDDALHVAITERMDPRISGEVVLGNPAAILDRSHDQLVLGRRASGLERMVGS 164  
QY 125 TAAALALSAQCPVAIVRPHRVPIGRDAMIYVADGSSDIG-VLLGAVMAEARLDSFVR 183  
DB 165 TSVAVAGMSAAPVVVIRASTPDPDGKCVAVAVGQSVGTAAVGFAPFADHRGCKL 224  
QY 184 VVTCRQSGVGTGDDVRAASLDRLARKQRPYRDPVQSAVHGEILLDTLAGRSYMMV 243  
DB 225 AVTVPGNDSEVHDEALKRLNEVVKPLANKPHVEVESTRVLGSPVALVDLSGVDLV 284  
QY 244 LSASDQEHVEQLVGAPGNVAVLOEAGCTLLV 273  
DB 285 IGMKHPILGWTAGVSRALMAHQSPLAT 314

RESULT 15  
ABU25937

ID ABU25937 standard; protein; 300 AA.

XX ABU25937;

DT 19-JUN-2003 (first entry)

XX Protein encoded by prokaryotic essential gene #11464.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Corynebacterium diphtheriae.

XX MO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002MO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948893.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX N-PSDB; ACA29807.

XX New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 53861; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than S. aureus, S. typhimurium,  
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 300 AA;

Query Match 12.8%; Score 180; DB 6; Length 300;  
Best Local Similarity 24.8%; Pred. No. 7.6e-10;  
Matches 73; Conservative 46; Mismatches 139; Indels 36; Gaps 9;

QY 10 IVVGIDSKPAVQALMAVDEAASRDIPRL-----LVA---IEPDDPGYAHGA 57  
DB 7 VVAVVDSSEASQNAVMAANTANKRGVPLRLAASYTMPQFLVAGVNPQELFDELQSE 66  
QY 58 ARCLAAENAVRAFAVFAADRPVKVEVEITQERPRVTLIRASAAALVCVGAIGVH 117  
DB 67 MDNIEAAR-----VVAHEVAP-DIKIGYVLAGSPIDMLDWSVDVTMIWGSRLG 119  
QY 118 RPERVGSTAALALSAQCPVAIVR-PHRVPIGRDAMIYVADGSSDIGVLLGAVMAEAR 176  
DB 120 SGVMKGSVGAAYVSHDCFPVVVRSNDNHTETNKGPVVVGVDSVSGRATFAPAEBA 179  
QY 177 LRDSP-VRVVTCRQSGVGTGDDVRAASLDRLARKQRPYRDPVQSAVHGEILLDT 221  
DB 180 ARGAKLVAIHTMDMDQVQASLAGLAAAOQSEWETIEKEQTLLKDRLLQPLERPRVEVM 239  
QY 222 AAHNGELLDYLAGRSYMMVVLASDQEHVE-QLVGAAGNVAVLOEAGCTLLV 274  
DB 240 VITDRPRVRLDECANNAQLLVGSHRGSGFRGMLIGSTRALLQSAFCPMVVV 293

Search completed: March 23, 2006, 05:10:40  
Job time : 77.1768 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using SW model

Run on: March 23, 2006, 05:11:18 ; Search time 11.1001 Seconds  
(without alignments)  
2418.401 Million cell updates/sec

Title: US-10-617-038-21

Perfect score: 1410  
Sequence: 1 MNOSHKEPSIVGIDGSKPA.....GNAVLOEAGCTLLVVGQOYL 279

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1410	100.0	279	2 C70942	hypothetical prote
2	431.5	30.6	268	2 G70645	hypothetical prote
3	416	29.5	272	2 G70572	hypothetical prote
4	257	18.2	295	2 C70759	hypothetical prote
5	224	15.9	294	2 T36949	conserved hypot
6	214	15.2	297	2 F70572	hypothetical prote
7	211.5	15.0	294	2 A70942	hypothetical prote
8	211	15.0	328	2 T36963	hypothetical prote
9	207	14.7	283	2 T36962	hypothetical prote
10	182.5	12.9	317	2 B70758	hypothetical prote
11	180.5	12.8	301	2 T37031	hypothetical prote
12	170.5	12.1	288	2 T37029	hypothetical prote
13	136	9.6	152	2 T36954	hypothetical prote
14	123	8.7	148	2 T36960	hypothetical prote
15	116	8.2	657	2 H70872	probable ctpd prot
16	114.5	8.1	2240	2 T37057	probable multi-dom
17	112.5	8.0	201	2 T37036	hypothetical prote
18	106.5	7.6	561	2 T36084	hypothetical prote
19	105	7.4	331	2 B87371	regulatory protein
20	104	7.4	154	2 A81272	hypothetical prote
21	103.5	7.3	250	2 B70502	hypothetical prote
22	102.5	7.3	1106	2 C87128	probable transcrip
23	101.5	7.2	143	2 A69233	conserved hypot
24	101.5	7.2	149	2 A69220	conserved hypot
25	101.5	7.2	402	2 B72502	probable ornithine
26	100.5	7.1	152	2 A81634	hypothetical prote
27	99.5	7.1	145	2 A86854	conserved hypot
28	99	7.0	442	2 C83256	hypothetical prote
29	99	7.0	3172	2 S22012	erythronolide synt

30	98.5	7.0	888	2 D87433	conserved hypot
31	98.5	7.0	664	2 T36410	probable polyketid
32	97.5	6.9	360	2 H70798	probable cation-tr
33	96.5	6.8	571	2 C75530	conserved hypot
34	96.5	6.8	2108	2 H70819	probable polyketid
35	96.5	6.8	3491	2 T43231	probable 6-deoxyer
36	96	6.8	150	2 G84340	hypothetical prote
37	96	6.8	490	2 B87023	UDP-N-acetylmutam
38	95	6.7	335	2 A71042	probable mevalonat
39	95	6.7	1763	2 T17465	rifamycin polyketi
40	94.5	6.7	157	2 A84402	hypothetical prote
41	94.5	6.7	274	2 A13643	citrate (pro-3S)-1
42	94.5	6.7	329	2 H70794	probable dnaq prot
43	94.5	6.7	354	2 D35255	probable trans-chl
44	94.5	6.7	413	2 A70795	hypothetical prote
45	94.5	6.7	463	2 A46449	hypothetical prote

ALIGNMENTS

```
RESULT 1
C70942
hypothetical protein Rv2028c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C:Accession: C70942
R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulemon, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: C70942
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-279 <COL>
A:Cross-references: UNIPROT:O53474; UNIPARC:UP100000D5EDC; GB:AL021899; GB:AL123456; NID
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv2028c

Query Match      100.0%; Score 1410; DB 2; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.4e-101;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MNOSHKEPSIVGIDGSKPAVQAALMAVDEAASRDIPRLILVAIEPDPGYAAHGAARK 60
      |||
Db      1 MNOSHKEPSIVGIDGSKPAVQAALMAVDEAASRDIPRLILVAIEPDPGYAAHGAARK 60

QY      61 LAAAEANVRYAFNVAEADRPVYVEVITTOERPPTSIRASAAALVCVCAIGVHHRRPE 120
      |||
Db      61 LAAAEANVRYAFNVAEADRPVYVEVITTOERPPTSIRASAAALVCVCAIGVHHRRPE 120

QY      121 RVGSTAAALASACCPAIVRPHRPVIGRDAAWIVAEADSSPDIGVLGAVMAEARLRDS 180
      |||
Db      121 RVGSTAAALASACCPAIVRPHRPVIGRDAAWIVAEADSSPDIGVLGAVMAEARLRDS 180

QY      181 PVRVVTGROSGVGDTGDDVNASIDRWLARWQPRYPDVVQSAAVHGEILDYLAGLGSVH 240
      |||
Db      181 PVRVVTGROSGVGDTGDDVNASIDRWLARWQPRYPDVVQSAAVHGEILDYLAGLGSVH 240

QY      241 MVTLSASDQEHVEQLVAPGNAYLOEAGCTLLVVGQOYL 279
      |||
Db      241 MVTLSASDQEHVEQLVAPGNAYLOEAGCTLLVVGQOYL 279

RESULT 2
G70645
hypothetical protein Rv3134c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
```

C:Accession: G70645  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajadaram, M.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.  
A>Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
A:Reference numbers: A70500; MUID:98295987; PMID:9634230  
A:Accession: G70645  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-268 <COL>  
A:Cross-references: UNIPROT: P95192; UNIPARC:UPI00000D603B; GB:Z83867; GB:AL123456; NID:9  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: Rv3134C

Query Match	30.6%	Score 431.5;	DB 2;	Length 266;
Best Local Similarity	40.4%	Pred. No. 4.3e-26;		
Matches	111;	Conservative	34;	Mismatches 105; Indels 25; Gaps 8;
Qy	1	MNOSHKEPSIVIGIDGSKRAVQAALMAVDKBAASRDIPRLTYLAIEPDDPGYAAHGAARK	60	
Dd	1	MSDRPPARAAYVGIDGSRAATHTAALMAVDENVNRDIPRLTYVIDPSQLSAAGGGQQ--	58	
Qy	61	LAAAEINAVAFITVEADRPVKYEVEITQRPYTSLIRASAAAALVCVAIGVHHFRPE	120	
Dd	59	-SARPAALHDHSRKVEATCGCPVKIETBYLCGRPLTKLMQBSRSMAMCIVSGVLGDHYRGR	117	
Qy	121	RVSSTAAALLASAOCFPAIVRPH--RVPIGDAAWIVEADSSDIGVLCGVAAEAERLR	178	
Dd	118	R-GSVAAATLASSALCPVAIVIHSPAPRAPATTSGQVAANVAEVDN----GVLTSHAFEEEARLR	172	
Qy	179	DSPRVVVTCRSGVGCDTGDDV-----RALDLRWLARQRYPDVYQSAAAHGELLDY	231	
Dd	173	GVPFLRAAVVH---AAETPDDVEQGSRLAHVHLISRLAHMTBLYPEVRDRAIAGGSACRH	229	
Qy	232	IAGGRSVHMVYLSSAQEHVEQLVGA--PGNAVL	264	
Dd	230	LAANAAPGQLFTV--ADSHSHAHEICGAIYQPCAAVL	261	

RESULT 3  
 G70572  
 hypothetical protein RV2624c - Mycobacterium tuberculosis (strain H37Rv)  
 C/Species: Mycobacterium tuberculosis  
 C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
 C/Accession: G70572  
 R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A/Authors: Squares, R.; Sultston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A/Reference number: A70500; MUID:98295987; PMID:9634230  
 A/Accession: G70572  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-272 <COL>  
 A/Cross-references: UNIPROT:O06188; UNIPARC:UPI00000D5EAL; GB:Z95387; GB:AL123456; NID:q  
 A/Experimental source: strain H37Rv  
 C/Genetics:  
 C/Gene: RV2624c

Query March	29.5%	Score 416;	DB 2;	Length 272;
Best Local Similarity	36.8%	Pred. No. 6.8e-25;		
Matches 98;	Conservative 46;	Mismatches 116;	Indels 6;	Gaps 2;
Oy	9	SIIVGIDGSKRAVQALMAVDEAASRDIPLELLYAIPEDDPGYAAHGCAARKLAANAV	68	
Db	11	TIIVGIDGSHAATITAAAGVDEAISRPAFLPLVSAVIKETHSPDPDYD--RDLAHARSL	67	
Oy	69	RYAETAVEADAPRKVEVEITQERPVTSILIRASAAALVCGAIGVHFRPRRGVSTAA	128	

```
Dbb      | : |||::||:~::~||:~::~|||  
68 RAAQAVAEAGLKVITETIPRGPAPVLVEASRDAMIVGSGVIGIRYASSIIISTATTE 127  
  
Oy     129 IALSACPAIYVRPHRVPIGRDAAMIIVEADSSDIGILLGAWMAEARLYSDSPVVVTTCR 188  
Db    128 LAEKHCPVAIVRSKIDOPASDINMIIVRMTDARPDEAVELEYAREAKLKQAPLIALAGR 187  
  
Oy     189 QSGVDGTGGDVDSLDRMLARQMQRYPDYRVOSAANYHGELLDYLAGEISVMNVLSASD 248  
Db    188 PELEWEIRP---GEFERRRVODMHHNRHPDYRVYPTTHTGIARFLADDERVKQLAIVIGGE 244  
  
Oy     249 QEHWEEVGARGNAVALQEAGCTLLLV 274  
Db    245 AGQLRLVPDSGHPPVFRAHECSVLVV 270
```

RESULT 4  
 C70759  
 Hypothetical protein RV2005C - Mycobacterium tuberculosis (strain H37Rv)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #ext\_change 09-Jul-2004  
 C:Accession: C70759  
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltywell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 353, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; NUID:98295987; PMID:9634230  
 A:Accession: C70759  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1285 <CDS>  
 A:Cross-references: UNIPROT:Q10851; UNIPARC:UPI0000013B62F; GB:274025; GB:AL123456; NID:g3  
 A:Experimental source: strain H37Rv  
 C:Genetics:  
 A:Gene: RV2005C

Query Match	18.2%;	Score 257;	DB 2;	Length 295;
Best Local Similarity	29.4%;	Pred. No. 1.4e-12;		
Matches	91;	Conservative 43;	Mismatches 126;	Indels 50; Gaps 9;
QY	1	NNQSHKPSIVVIGIGISKEPVAQALMAVDEAASRDIPILRLVLAIEPD-----	47	
DB	1	MSKPKRGICGVVGVGVSLESDAACGATDAARNRILPLVYVHVYVNDVATWPMRPYETW 60		
QY	48	----DPCYAAHGAARKLAALAAENAVRYAFTVAEADRPVKVEVEITQERPTSLIPASA 102		
DB	61	GVWQDEG-----RQIVV--NAVKLAKAAY-GADRKLSYKSELVFSYTPVPTWVISN 109		
QY	103	AAALVCVGAIGVHHFRPERVGSSTAALALSAQCPLVLR-----PHRVP--IGRDA 152		
DB	110	EAEWVLSSGGGALARGLLGSVSSSLVRACCPVAVHISDDAVIDPDQHPALVVLGIDGS 169		
QY	153	WI-----VEADGSSDIDGLGAVMAEARLDRSPRYVUTCROSGVGDTCDDVPAASLDRL 207		
DB	170	PVSELATVAAPFBAARRGVELLVAIVHAM-----SDVEVELPGIDFSAVQGEAELSIAERL 224		
QY	208	ARMQRRYPDPVQSAAVHGELLIDVLAGLGRSYHMVVLLSASDOEHV--QLVGAPGNVAVLQE 266		
DB	225	AGWGRYRPDPVPSRVVVCDRPARKLVQKGSASQILVVVGSHGRCGLTGMLIGSVSNVATLHA 284		
QY	267	AGCTLLVVQ 276		
DB	285	ARVPVIVARQ 294		

RESULT 5  
T36949  
conserved hypothetical protein SCJ1.16c - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #ext\_change 09-Jul-2004  
C:Accession: T36949





A;Residues: 1-301 <MUR>  
A;Cross-references: UNIPROT:Q9RI46; UNIPARC:UPI00000DB3E6; EMBL:AL109989; PIDN:CAB53424.  
A;Experimental source: strain A3(2)  
C;Genetics:  
A;Gene: SCOEDB:SCJ12.12C

Query Match 12.8%; Score 180.5; DB 2; Length 301;  
Best Local Similarity 25.4%; Pred. No. 1.1e-06;  
Matches 76; Conservative 40; Mismatches 136; Indels 53; Gaps 10;

QY 9 SIVVIGDSKPAVQALMAVDEAASRDIPRLILVAIEPDDPGYAAHGAARKLAARNAV 68  
D 4 TITGLGSPSPRAAEWAAREGLTRVPRILVHWQP-----VPEPMAQAPLLGAEHQ 58  
QY 69 RYATVAEADRAPRKVE---VEITQER---PTSLIRASAAALVCGALGVHFRPER 121  
D 59 HMTKRIPRDTREGRLRHPRGVETTEQATGNPADALLAGTIDABILVGSALSGLGFL 118  
QY 122 VGSTAAALASAOCPVAIVR-----PHRVPIGRDAW-----IIVEADGSSDYG 165  
D 119 VGSVGSVIATERTPVLIVRAGEQAADENLKDPGIP---SAATGFRVTVVGLDTGSPDE 175  
QY 166 VLGAVMAREALRDSPPRVVTCRQSGVGT-----GDDVRASLDRLMAR-----WQP 212  
D 176 AVLSFAFEAEARRRRAPLTAVAWNLPSSTYTSIAAGFPREELARAQALGALPWRRE 235  
QY 213 RYPVRVQSAVAHGEILDYLAGRSVHMVYLSASDOE-----HYEQLVGARGNAVLOEA 267  
D 236 KYPVEVETETCRSLSPAEHLIDAAADASLVVGGRRIRKSPGVH---IGAVIAVHMHA 291  
QY 268 GCTLLVV 274  
D 292 TTPVAVV 298

# RESULT 12

T37029  
hypothetical protein SCJ12.10C - Streptomyces coelicolor  
C;Species: Streptomyces coelicolor  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C;Accession: T37029  
R;Murphy, L.; Harris, D.; Thomson, N.R.; Parhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, August 1999  
A;Reference number: Z21619  
A;Accession: T37029  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-288 <MUR>  
A;Cross-references: UNIPROT:Q9RI48; UNIPARC:UPI00000DB3E4; EMBL:AL109989; PIDN:CAB53422.  
A;Experimental source: strain A3(2)  
C;Genetics:  
A;Gene: SCOEDB:SCJ12.10C

Query Match 12.1%; Score 170.5; DB 2; Length 288;  
Best Local Similarity 25.2%; Pred. No. 6.3e-06;  
Matches 76; Conservative 42; Mismatches 125; Indels 59; Gaps 11;

QY 10 IVVIGDSKPAVQALMAVDEAASRDIPRLILVAIEPDDPGYAAHGAARKL-----A 62  
D 5 MVVVDVDESSIGAVDMAADEALHEVLRIVHAYRMD---RYEGASLARELKKPSGHV 60  
QY 63 AAEAVRATFAVEADRPVAVVEITQERPVTSILIRASAAALVCGALGVHFRPERV 122  
D 61 TTDTLAVATRRARRHFDLAVTTTEATIEEPYILLRARRASAVIILSTRGRGELAGLL 120  
QY 123 GSTAAALASAOCPVAIVR-PH-----RVPIG-RDAAVIVE--ADGSSDYGVL 167  
D 121 GSVSLVATWSDCPVTVTRGSHDDRAAGGRGRIVGVADAPRAAVAFACEBARRGAA 180  
QY 168 LGAVMA-----EAKRLDSPPRVVTCR-----QSGVGTGDDVRASLDRLMARQOP 212  
D 181 LDAVAVRACPTHTDVDPHFLTAGTBERLHEERAKLEBALADADAVR-----LRRRTA 234

QY 213 RYEDVRVQSAVAHGEILDYLAGRSVHMVYLSASDOEHVEQLVGARGNALVDEAGCTLL 272  
D 235 EGPSRSVLSAASH-----EADLLVGGRRRPGFGHRLGRVAHTLLHRSACPVA 282

QY 273 VV 274  
D 283 VV 284

# RESULT 13

T36954  
hypothetical protein SCJ1.21 - Streptomyces coelicolor  
C;Species: Streptomyces coelicolor  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C;Accession: T36954  
R;Seeger, K.J.; Harris, D.; Thomson, N.R.; Parhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, September 1999  
A;Reference number: Z21607  
A;Accession: T36954  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-152 <SEE>  
A;Cross-references: UNIPROT:Q9RI23; UNIPARC:UPI00000DB393; EMBL:AL109962; PIDN:CAB53139..  
A;Experimental source: strain A3(2)  
C;Genetics:  
A;Gene: SCOEDB:SCJ1.21  
A;Superfamily: Escherichia coli ybdQ protein

Query Match 9.6%; Score 136; DB 2; Length 152;  
Best Local Similarity 29.8%; Pred. No. 0.0014;  
Matches 45; Conservative 17; Mismatches 71; Indels 18; Gaps 2;

QY 1 MNOSHKRPSIVGIGDSKPAVQALMA-----VDEAASRDIPRLILVAIEPDDPG 50  
D 1 MOPSSEPTARVYVGVDSPPSSYALRMADRYARAVGVVEAVHWDTSAVGFACPAIDPD 60  
QY 51 YAAHGAARKLAAENAVRYAFTAVEADRPVKEVEITQERPVTSILIRASAAALVCGV 110  
D 61 FDLGQARERFAELE-----ATFGERPPGLKETILVGDPEETLIRASGAEILLVVG 112  
QY 111 ALGVHFRPERVGSTAAALASAOCPVAIVR 141  
D 113 RRGGAFAARMGLSVSGRCAQHAACPVVVR 143

# RESULT 14

T36960  
hypothetical protein SCJ1.27 - Streptomyces coelicolor  
C;Species: Streptomyces coelicolor  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C;Accession: T36960  
R;Seeger, K.J.; Harris, D.; Thomson, N.R.; Parhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, September 1999  
A;Reference number: Z21607  
A;Accession: T36960  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-148 <SEE>  
A;Cross-references: UNIPROT:Q9RI17; UNIPARC:UPI00000DB399; EMBL:AL109962; PIDN:CAB53145..  
A;Experimental source: strain A3(2)  
C;Genetics:  
A;Gene: SCOEDB:SCJ1.27

Query Match 8.7%; Score 123; DB 2; Length 148;  
Best Local Similarity 29.7%; Pred. No. 0.014;  
Matches 43; Conservative 18; Mismatches 66; Indels 18; Gaps 2;

QY 97 LIRASAAALVCGALGVHFRPERVGSTAAALASAOCPVAIVRPHRVPIGRDAAVIV 156  
D 2 LVQAALSAAVVGVGHGRPHFGALIGSLATGLAAHSECTVVVVGFEENRAGPVALGVDG 61  
QY 157 EADGSSDYG-----LLGAVMAEARLDSFVRVVTCTRQSGVGTGDDVRA 201

Db 62 SAAGAAAVRLPGDFRACGHHRRARLDGLERADA---SPRRTEPYACEFGMLAEGR 118

Qy 202 SLDRWLARWQRYPDVVRQSAVAHG 226

Db 119 LLAETLAGMCARYPEVRVERSTVHG 143

## RESULT 15

H70872

probable cypD protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C&gt;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004

C:Accession: H70872

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; PMID:98295987; PMID:9634230

A:Accession: H70872

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-657 &lt;COL&gt;

A:Cross-references: UNIPROT:O53160, UNIPARC:UPI0000128642, GB:AL021184, GB:AL123456, NIT

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: cypD

C:Superfamily: Enterococcus copper-transporting ATPase codb; ATPase nucleotide-binding d

F;25-394/Domain: ATPase transduction domain homology &lt;ATP&gt;

F;465-608/Domain: ATPase nucleotide-binding domain homology &lt;ATN&gt;

## Query Match

8.2%; Score 116; DB 2; Length 657;

Best Local Similarity 25.7%; Pred. No. 0.25;

Matches 69; Conservative 32; Mismatches 101; Indels 66; Gaps 13;

Qy 27 AVDEAA--SRDIPRLIYAIPEDDPGYAAHGAARKLAAMNAVYAF---TAVEAADRP 81

Db 345 ALDKTGTLTRGIP--RLASVAFIDPNVVDARRLLQLAAAEQSSSEHPLGRIVAEARRRG 402

Qy 82 VKVE-----VEITQER-----PVTSIRASAAAL-----V 107

Db 403 IAIIPKADFPRAVPCGVHALVGNDFEIASPQSYRGAPLAEPLASAGATPAIVLLDGV 462

Qy 108 CVGAIGV-IHFRPERVGTAAALALSAQCPVAIVPRHVPICRDAMI VVEADGSSDIGN 166

Db 463 AIGVIGLTQDRPDVAESVAAMALTAAPVLLTGDN----GR-AAWRVARNNGITDVRA 517

Qy 167 LIGAVMABRLRDSPVRVVTCRQSG-----VGPTGDVRASLDRWLARWQRYPDVVRQ 220

Db 518 AL-----LPEQKVEVVNRLQAGHQVLLVGDGVNDAPAMAAARAAMVAMGAGADLTQ 569

Qy 221 SA---AVHGEL--LDVLAGRSVHMV 243

Db 570 TADGVTTRDELHTIPTITGLARQARRV 597

Search completed: March 23, 2006, 05:26:27

Job time : 12.1001 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model  
Run on: March 23, 2006, 04:56:43 ; Search time 71.3577 Seconds  
(without alignments)  
2758.529 Million cell updates/sec

Title: US-10-617-038-21  
Perfect score: 1410  
Sequence: 1 MNOSHKPESIVVGIDSKPA.....GNAVLQEAAGCTLLVVGQOYL 279

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues  
Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_05.80:\*  
1: uniprot\_sprotc:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1410	100.0	279	Q7T244_MYCBO	Q7T244 mycobacteri
2	1410	100.0	279	O53474_MYCTU	O53474 mycobacteri
3	431.5	30.6	268	Q7TX68_MYCBO	Q7TX68 mycobacteri
4	431.5	30.6	268	P95192_MYCTU	P95192 mycobacteri
5	416	29.5	272	Q7TY74_MYCBO	Q7TY74 mycobacteri
6	416	29.5	272	O06188_MYCTU	O06188 mycobacteri
7	374.5	26.6	252	Q73V37_MYCPA	Q73V37 mycobacteri
8	257	18.2	295	Y2005_MYCTU	Y2005 mycobacteri
9	257	18.2	295	Y2028_MYCBO	Y2028 mycobacteri
10	238	16.9	281	Q7WZ45_9ACTO	Q7WZ45 nonomuraea
11	227	16.1	294	Q73Z49_MYCPA	Q73Z49 mycobacteri
12	224	15.9	294	Q9R128_STRCO	Q9R128 streptomyc
13	214.5	15.2	294	Q73Z62_MYCPA	Q73Z62 mycobacteri
14	214	15.2	297	Q7TY75_MYCBO	Q7TY75 mycobacteri
15	214	15.2	297	O06189_MYCTU	O06189 mycobacteri
16	213	15.1	343	Q4NL54_9M1CC	Q4NL54 arthrobacte
17	211.5	15.0	294	Q7TZ45_MYCBO	Q7TZ45 mycobacteri
18	211.5	15.0	294	O53472_MYCTU	O53472 mycobacteri
19	211	15.0	328	Q9R1Y4_STRCO	Q9R1Y4 streptomyc
20	208	14.8	295	O82NK3_STRCO	O82NK3 streptomyc
21	207	14.7	283	Q9R1Y5_STRCO	Q9R1Y5 streptomyc
22	207	14.7	296	O5YV67_NOCFA	O5YV67 nocardia fa
23	207	14.7	301	Q8NLJ9_CORGL	Q8NLJ9 streptobacte
24	206.5	14.6	290	Q82NM7_STRCO	Q82NM7 streptomyc
25	206	14.6	345	O5YV44_NOCFA	O5YV44 nocardia fa
26	203.5	14.4	293	Q73Z61_MYCPA	Q73Z61 mycobacteri
27	203.5	14.4	296	O5YTS4_NOCFA	O5YTS4 nocardia fa
28	202	14.3	318	Q8FLR8_CORER	Q8FLR8 corynebacte
29	196	13.9	293	Q5YV67_NOCFA	Q5YV67 nocardia fa
30	190	13.5	284	O6A8Y6_PROAC	O6A8Y6 propionibac
31	188	13.3	303	Q4Y55_CORJX	Q4Y55 corynebacte

32	183.5	13.0	300	Q9K4H8_STRCO	Q9K4H8 streptomyc
33	182.5	12.9	317	Y11996_MYCTU	Y11996 mycobacteri
34	182.5	12.9	317	Y2019_MYCBO	Y2019 mycobacteri
35	180.5	12.8	301	Q9R146_STRCO	Q9R146 streptomyc
36	180	12.8	300	O6NEJ7_COROI	O6NEJ7 corynebacte
37	171	12.1	312	Q9K4L5_STRCO	Q9K4L5 streptomyc
38	170.5	12.1	288	Q9R148_STRCO	Q9R148 streptomyc
39	167.5	11.9	144	Q4NDF1_9M1CC	Q4NDF1 arthrobacte
40	157	11.1	321	O8G872_BIFLO	O8G872 bifidobacte
41	154	10.9	345	O8G3U1_BIFLO	O8G3U1 bifidobacte
42	148	10.5	151	Q4NGW3_9M1CC	Q4NGW3 arthrobacte
43	146.5	10.4	284	Q4NL45_9M1CC	Q4NL45 arthrobacte
44	139.5	9.9	289	Q4NE71_9M1CC	Q4NE71 arthrobacte
45	138	9.8	159	Q4NDE9_9M1CC	Q4NDE9 arthrobacte

ALIGNMENTS

RESULT 1	Q7T244_MYCBO	PRELIMINARY:	PRT:	279 AA.
ID	Q7T244_MYCBO			
AC	Q7T244			
DT	01-OCT-2003 (Tremblrel. 25, Created)			
DT	01-OCT-2003 (Tremblrel. 25, Last sequence update)			
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)			
DE	Hypothetical protein MB2053c.			
GN	OrderedLocustNames=MB2053c;			
OS	Mycobacterium bovis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacteriineae; Mycobacteriaceae; Mycobacterium;			
OC	Mycobacterium tuberculosis complex.			
NC	NCBI_TaxID=1765;			
RM	[1]			
RP	NCBI_EOTIDR SEQUENCE.			
RC	STRAIN=AF2122/97;			
RX	MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;			
RA	Garnier T., Eigelmeier K., Camus J.-C., Medina N., Mansoor H.,			
RA	Pryor M., Duthey S., Gordin S., Lacroix C., Monsemp C., Simon S.,			
RA	Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,			
RA	Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewison R.G.,			
RT	"The complete genome sequence of Mycobacterium bovis."			
RL	Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).			
DR	EMBL; BX248341; CAD96906.1; -; Genomic DNA.			
DR	GO; GO:0006950; P:response to stress; IEA.			
DR	InterPro; IPR006015; Usp.			
DR	InterPro; IPR006016; UspA.			
DR	Pfam; PF00582; Usp; 1.			
DR	PRINTS; PR01438; UNWRSLSTRESS.			
KW	Complete proteome; Hypothetical protein.			
SQ	SEQUENCE 279 AA; 29441 MW; 52F240F9FA466AC9 CRC64;			
Query Match	100.0%;	Score 1410;	DB 2;	Length 279;
Best Local Similarity	100.0%;	Pred. No. 7.5e-102;		
Matches 279;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MNOSHKPESIVVGIDSKPAVOALNAVDEASRDIPRLLYAIEPDGPYAAHGAARK	60	
DB	1	MNOSHKPESIVVGIDSKPAVOALNAVDEASRDIPRLLYAIEPDGPYAAHGAARK	60	
QY	61	LAALAAVRAFAFAVEADRPVKEVEITQERPVTSILIRASAAALVCVAGIYVHFRPE	120	
DB	61	LAALAAVRAFAFAVEADRPVKEVEITQERPVTSILIRASAAALVCVAGIYVHFRPE	120	
QY	121	RVGSTAALALSAQCPAIVRPHRPVIGRANATIVVADSSSDIGVLLGVMAEARLDS	180	
DB	121	RVGSTAALALSAQCPAIVRPHRPVIGRANATIVVADSSSDIGVLLGVMAEARLDS	180	
QY	181	PVRVVTGROSGVGTGDDVVASLDRMLARMQRPYDPVRVOSAAVHGEILDYLAGISGVH	240	
DB	181	PVRVVTGROSGVGTGDDVVASLDRMLARMQRPYDPVRVOSAAVHGEILDYLAGISGVH	240	
QY	241	MVVLASDQERHVEQLVGAPGNAVLQEAAGCTLLVVGQOYL	279	

Db 241 MVLASDQEHVEQLVGAPGNAVLQEGCTLLVVGQYL 279

|||||

RESULT 2

053474 MYCTU PRELIMINARY; PRT; 279 AA.

ID 053474 MYCTU PRELIMINARY; PRT; 279 AA.

AC 053474; Q7D7L5;

DT 01-JUN-1998 (TrEMBLrel. 06, Created)

DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)

DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)

DE Hypothetical protein (universal stress protein family).

GN OrderedLocustNames=MT2087, Rv2028c;

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;

OC Mycobacterium tuberculosis complex.

NCBI\_TaxID=1773;

OX [1];

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=H37Rv; PubMed=9634230; DOI=10.1038/31159;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C.M.,

RA Harris D.E., Gordon S.V., Eigmeier K., Gae S., Barry C.E. III,

RA Tekata F., Badcock K., Bauman D., Brown D., Chillingworth T.,

RA Connor R., Davies R.W., Devlin K., Felwell T., Gentles S., Hamlin N.,

RA Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S.,

RA Murphy L.D., Oliver S., Osborne K., Quail M.A., Rajandream M.A.,

RA Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,

RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the

RT complete genome sequence."

RL Nature 393:537-544(1998).

RN [2]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=CDC 1551 / Oshkosh; PubMed=12218035;

RX MEDLINE=22206494; PubMed=12218035;

RX DOI=10.1128/JB.184.19.5479-5490.2002;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

RA Peterson J.D., DeBoy R.T., Dodson R.J., Gwinn M.L., Haft D.H.,

RA Hickey E.K., Kolonay J.F., Nelson W.C., Umayam L.A., Esmailova M.D.,

RA Salzberg S.L., Delcher A.L., Uterback T.R., Weidman J.F., Kouri H.M.,

RA Gill J., Mikha A., Bishai W., Jacobs W.R. Jr., Venter J.C.,

RA Fraser C.M.;

RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and

RT laboratory strains."

RL J. Bacteriol. 184:5479-5490(2002).

DR EMBL: BX842578; CAI17242.1; -; Genomic DNA.

DR EMBL: AE000516; AKK46366.1; -; Genomic DNA.

DR FIC: C70942; C70942.

DR TIGR: MT2087; -.

DR Tuberculins; Rv2028c; -.

DR GO: GO:0006950; P:response to stress; IEA.

DR Interpro: IPR006015; Usp.

DR Interpro: IPR006016; UspA.

DR Pfam: PF00582; Usp; 1.

DR PRINTS: PRO1438; UNVRSLSTRESS.

KW Complete proteome; Hypothetical protein.

SQ SEQUENCE 279 AA; 29441 MW; 52F240F9FE466AC9 CRC64;

Query Match 100.0%; Score 1410; DB 2; Length 279;

Best Local Similarity 100.0%; Pred. No. 7, 5e-102; Indels 0; Gaps 0;

Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNQSHKPPSTIVGIDGSKPAVQALMAVDEASRDIPRLILYALIEPDDPGYAAHGAARK 60

DB 1 MNQSHKPPSTIVGIDGSKPAVQALMAVDEASRDIPRLILYALIEPDDPGYAAHGAARK 60

QY 61 LAAAEANVRYAFTAVEAADRPVKEVEITQERPVTSIRASAAALVCVAGIYGHFRPE 120

DB 61 LAAAEANVRYAFTAVEAADRPVKEVEITQERPVTSIRASAAALVCVAGIYGHFRPE 120

QY 121 RVGSTAALALSLAQCPVAIVRPHRPVPIGRDAAMIVVEADGSSDGLVGLGAVMAEARLDS 180

Db 241 MVLASDQEHVEQLVGAPGNAVLQEGCTLLVVGQYL 279

|||||

RESULT 3

ID 07TX68 MYCBO PRELIMINARY; PRT; 268 AA.

AC 07TX68;

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Hypothetical protein M03158c.

GN OrderedLocustNames=M03158c;

OS Mycobacterium bovis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;

OC Mycobacterium tuberculosis complex.

NCBI\_TaxID=1765;

OX [1];

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=AF2122/97; PubMed=12788972; DOI=10.1073/pnas.1130426100;

RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,

RA Pryor M., Duthey S., Grondin S., Lacroix C., Monsemp C., Simon S.,

RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,

RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;

RT "The complete genome sequence of Mycobacterium bovis."

RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).

DR EMBL: BX248345; CAD95250.1; -; Genomic DNA.

DR GO: GO:0006950; P:response to stress; IEA.

DR Interpro: IPR006015; Usp.

DR Interpro: IPR006016; UspA.

DR Pfam: PF00582; Usp; 1.

DR PRINTS: PRO1438; UNVRSLSTRESS.

KW Complete proteome; Hypothetical protein.

SQ SEQUENCE 268 AA; 28008 MW; EE2CB4F9C9D602AE CRC64;

Query Match 30.6%; Score 431.5; DB 2; Length 268;

Best Local Similarity 40.4%; Pred. No. 1, 6e-25;

Matches 111; Conservative 34; Mismatches 105; Indels 25; Gaps 8;

QY 1 MNQSHKPPSTIVGIDGSKPAVQALMAVDEASRDIPRLILYALIEPDDPGYAAHGAARK 60

DB 1 MNDPRPARAVVIGIDGSKPAVQALMAVDEAVVRDIPRLILYALIEPDDPGYAAHGAARK 58

QY 61 LAAAEANVRYAFTAVEAADRPVKEVEITQERPVTSIRASAAALVCVAGIYGHFRPE 120

DB 59 -SARPAALHDASKKVEITQGPVKEITQERPVTSIRASAAALVCVAGIYGHFRPE 117

QY 121 RVGSTAALALSLAQCPVAIVRPHRPVPIGRDAAMIVVEADGSSDGLVGLGAVMAEARL 178

DB 118 R-GSVATLALGSLCPVAIVRPHRPVPIGRDAAMIVVEADGSSDGLVGLGAVMAEARL 172

QY 179 DSEVRVYTCRQSGVGTGDDV-----RASLDRLMLARQPRPDVVRVQSAVHGLLDY 231

DB 173 GVPLRAVAVH---AAETPDDVEGSRILAHVHSRRILAHVTRLYBEVRVRAIAGSGACRH 229

QY 232 LAGLGSRVHVAVLASDQEHVEQLVGA--PGNAV, 264

DB 230 LAAAKRGQLFV---ADSHSAHELCAVYDPCAVL 261

RESULT 4

ID P95192 MYCTU PRELIMINARY; PRT; 268 AA.

AC P95192; Q7D624;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
 DE Hypothetical protein (Universal stress protein family).  
 GN OrderedLocustNames=MT3220, Rv3134c;  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;  
 OC Mycobacterium tuberculosis complex.  
 NC NCB1\_TaxID=1773;  
 RX NCBI [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230; DOI=10.1038/11159;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C.M.,  
 Harris D.E., Gordon S.V., Eigmeier K., Gae S., Barry C.E. III,  
 Tekala F., Badcock K., Baaham D., Brown D., Chillingworth T.,  
 Connor R., Davies R.M., Devlin K., Feltwell T., Gentles S., Hamlin N.,  
 Holtroyd S., Hornby T., Jagsels K., Krogh A., McLean J., Moule S.,  
 Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,  
 Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 Sulston J.E., Taylor K., Whitehead S., Barrett B.G.,  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RX MEDLINE=22206494; PubMed=12218036;  
 RX DOI=10.1128/JB.184.19.5479-5490.2002;  
 RA Pleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 Peterson J.D., DeBoy R.T., Dodson R.J., Gwinn L.A., Haft D.H.,  
 Hickey E.K., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D.,  
 Salzberg S.L., Delcher A., Utermbeck T.R., Weisman J.F., Kouri H.M.,  
 Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
 laboratory strains."  
 RL J. Bacteriol. 184:5479-5490(2002).  
 DR EMBL; BX842582; CAB06280.1; -; Genomic\_DNA.  
 DR EMBL; AE000516; AAK47558.1; -; Genomic\_DNA.  
 DR PIR; G70645; G70645.  
 DR TIGR; MT3220; -;  
 DR Tuberculist; Rv3134c; -;  
 DR GO; GO:0006950; P:response to stress; IEA.  
 DR InterPro; IPR006015; Unsp.  
 DR InterPro; IPR006016; Unsp.  
 DR Pfam; PF00582; Unsp; 1.  
 DR PRINTS; PR01438; UNRS1STRESS.  
 KW Complete proteome; Hypothetical protein.  
 SO SEQUENCE 268 AA; 28008 MW; EE2CB4F9C9D602AE CRC64;

Query Match 30.6%; Score 431.5; DB 2; Length 268;  
 Best Local Similarity 40.4%; Pred. No. 1.6e-25;  
 Matches 111; Conservative 34; Mismatches 105; Indels 25; Gaps 8;

QY 1 MNQSHKPSIVIGIDGKPAVOALMAVDEAASRDIPRLYLAIPEDDPGYAAAGAAARK 60  
 DB 1 MSDDRPAVVVVGIDGSRPAATMAAMAVDEAVNRDIPRLYLIVIDPSLSAAGCGGCG-- 58  
 QY 61 LAAENAVRYAFTVVEADRPVYKVEITQERPVTSILRASAAALVVCYGAIGVHPRPE 120  
 DB 59 -SARPAALHDSRKVEATGCGVKTETELCGRPFLTKMQESRSAMLCVSGVGLDHVGR 117  
 QY 121 RVGSTAALALSAOCQPAIVPRH--RVDIGDAWIVVEADGSSDIDGLVAGVAEAARLR 178  
 DB 118 R-GSVAAATLQASALCPVAIVHSPAPRATTSQVSAVVAEVDN---GVLLHAFEEARLR 172  
 QY 179 DSPRVVTCRSGVGDIGDDV-----RASIDRLMAWQRRYPDVRYQSAVHGGELIDY 231  
 DB 173 GVPRAVAVH--AAETPDVDEOGSRLLAHVHLSRLAHMTLYPEVRVDRALVAGGSACRH 229

QY 232 LAGRSVHVNVISASDQEHVEQLVGA--PQNAV 264  
 DB 230 LAANAKRGOLFV---ADSHAEHLGAYOPGCAVL 261

RESULT 5

ID Q7Y74\_MYCBO PRELIMINARY; PRT; 272 AA.  
 AC Q7Y74;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Hypothetical protein Mb2657c.  
 GN OrderedLocustNames=Mb2657c;  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;  
 OC Mycobacterium tuberculosis complex.  
 NC NCB1\_TaxID=1765;  
 RX NCBI [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=AF2122/97;  
 RX MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;  
 RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,  
 Pryor M., Dutfoy S., Grondin S., Lacroix C., Monempe C., Simon S.,  
 Harris B., Alkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 RA Parkhill J., Barrett B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
 RT "The complete genome sequence of Mycobacterium bovis."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
 DR EMBL; BX248343; CAD94842.1; -; Genomic\_DNA.  
 DR GO; GO:0006950; P:response to stress; IEA.  
 DR InterPro; IPR006015; Unsp.  
 DR InterPro; IPR006016; Unsp.  
 DR Pfam; PF00582; Unsp; 2.  
 DR PRINTS; PR01438; UNRS1STRESS.  
 KW Complete proteome.  
 SO SEQUENCE 272 AA; 29400 MW; 3EE3A17B3C5720DB CRC64;

Query Match 29.5%; Score 416; DB 2; Length 272;  
 Best Local Similarity 36.8%; Pred. No. 2.6e-24;  
 Matches 98; Conservative 46; Mismatches 116; Indels 26; Gaps 2;

QY 9 SIVGIDGSKPAVOALMAVDEAASRDIPRLYLAIPEDDPGYAAAGAAARKLAAENAV 68  
 DB 11 TIIVIGDSNAITRALMGVDEAISRAVPLRVIVIKPTSPDDVD---RDLHAERSL 67  
 QY 69 RYAFVAAADRPVYKVEITQERPVTSILRASAAALVVCYGAIGVHPRPVGSTAA 128  
 DB 68 REAQSAAVEAAGKVKITETDIPRGAPGLVVEASRDAMICVSGVIGRYASSTIGSTATE 127  
 QY 129 LALSACQPAIVPRHVRPIGRDAWIVVEADGSSDIDGLVAGVAEAARLDSPVVTCR 188  
 DB 128 LAERGAHCPVAVWMSKVDOPASDINWIVMTDADNEAVLEIYAREKKLQAPLALGR 187  
 QY 189 QSGVGDIGDDVRSIDRLMAWQRRYPDVRYQSAVHGGELIDYLAGRSVHVNVISASD 248  
 DB 188 PEIRARELPD---GEFERVQDMHNRHVDVYVPIITHTGTARFLADDERVQLAVIGGE 244  
 QY 249 QEHVEQLVGAQNAVLOEAGCTLLIV 274  
 DB 245 AGQLARLVGPSGHVFRRHAECSVLVV 270

RESULT 6

ID Q06188\_MYCTU PRELIMINARY; PRT; 272 AA.  
 AC Q06188; Q7D6V6;  
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
 DE Hypothetical protein (Universal stress protein family).  
 GN OrderedLocustNames=MT2699, Rv2624c;  
 OS Mycobacterium tuberculosis.

```

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriineae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium tuberculosis complex.
OX NCBI_TaxID=1773;
RN
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=634230; DOI=10.1038/31159;
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C.M.,
RA Harris D.E., Gordon S.V., Eiglmeyer K., Gae S., Barry C.E. III,
RA Tekala F., Badcock K., Basham D., Brown D., Chillingworth T.,
RA Connor R., Davies R.M., Devlin K., Fellwell T., Gentles S., Hamlin N.,
RA Holroyd S., Hornsby T., Jagsels K., Krogan A., McLean J., Moule S.,
RA Murphy L.D., Oliver S., Osborne K., Quail M.A., Rajandream M.A.,
RA Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultison J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RX DOI=10.1128/JB.184.19.5479-5490.2002;
RA Fleisichmann R.D., Alland D., Eisen J.A., Carpenter J., White O.,
RA Peterson J.D., DeBoy R.T., Dodson R.J., Gwinn M.L., Haft D.H.,
RA Hickey E.K., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D.,
RA Salzberg S.L., Delcher A.L., Uterback T.R., Weidman J.F., Kouri H.M.,
RA Gill J., Mikula A., Bhat N., Jacobs W.R. Jr., Venter J.C.,
RA Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
DR EMBL: BX842580; CAB08618.1; -; Genomic DNA.
DR EMBL: AB000516; AAK47015.1; -; Genomic DNA.
DR FTR: G70572; G70572.
DR TIGR: MT2699; -.
DR TubercuList; Rv2624c; -.
DR GO: GO:0006950; P:response to stress; IEA.
DR InterPro: IPR006015; Usp.
DR InterPro: IPR006016; UspA.
DR Pfam: PF00582; Usp; 2.
DR PRINTS: PRO1438; UNVRS1STRESS.
KM Complete proteome.
SQ SEQUENCE 272 AA; 29400 MW; 3EE3A17B3C5720DB CRC64;

Query Match 29.5%; Score 416; DB 2; Length 272;
Best Local Similarity 36.8%; Pred. No. 2,6e-24;
Matches 98; Conservative 46; Mismatches 116; Indels 6; Gaps 2;

QY 9 SIIVGIDSGKPAVQALMAVDEASRDIPRLVLAIEPDDPGVAAHGAARKLAENAV 68
DB 11 TIIVGIDSGKPAVQALMAVDEASRDIPRLVLAIEPDDPGVAAHGAARKLAENAV 67
QY 69 RYAFVAVAEADRPVKEVEITQERPTSLIRASAAALVCVGAIGVHFRPERVSTAA 128
DB 68 REAQAQVAVAEAGKLVKRTIDIPRGPAQVLEASRDAMICVSGVIGRYASSILGSTATE 127
QY 129 LAASQCPVAVIYPRHPIPIGRDAWIVYVADSSDGLVLCGAAWAEKRLDSPPRYVTCR 188
DB 128 LAKRAHCPVAVWMSKVDQAPASDINWIVMTDADNEAVLEYAAREKRLQAPLALGGR 187
QY 189 OSGVGTGDDVRAASLDRLMARMPRPYDPVRVQSAAVGELLDYLAGRSGHNVVLSASD 248
DB 188 PEELIRELPD--GEFERVVDMMHRRHPDVAVYPTTTGTGARFLADDERVQLAVIGGGE 244
QY 249 QEHVEQIVGAPGNAVLQEAAGCTLLV 274
DB 245 AGQLARLVGSGHPVFHAECSVLV 270

```

RESULT 7  
073V37\_MYCPA

```

ID 073V37_MYCPA PRELIMINARY; PRT; 252 AA.
AC 073V37;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=MAP3179c;
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriineae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium avium complex (MAC).
OX NCBI_TaxID=1770;
RN
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K10;
RA Li L., Bannantine J., Zhang Q., Amosn A., Alt D., Kapur V.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ database.
DR EMBL: AB017238; AA505727.1; -; Genomic DNA.
DR GO: GO:0006950; P:response to stress; IEA.
DR InterPro: IPR006016; UspA.
DR Pfam: PF00582; Usp; 1.
KM Complete proteome; Hypothetical protein.
SQ SEQUENCE 252 AA; 26447 MW; FFDAB636B3BA80B1 CRC64;

Query Match 26.6%; Score 374.5; DB 2; Length 252;
Best Local Similarity 39.4%; Pred. No. 4.1e-21;
Matches 106; Conservative 41; Mismatches 93; Indels 29; Gaps 12;

QY 1 MNQSHKPSIVGIDSGKPAVQALMAVDEASRDIPRLVLAIEPDDPGVAAHGAAR 59
DB 1 MSTRVAPPAIVGVDSRRAMHAAVWAIDAVGRDIPRLVLYVIDHGAAG--GHGPDTR 58
QY 60 KLAANAENAVYAFVAAEADRPVKEVEITQERPTSLIRASAAALVCVGAIGVHFR-- 117
DB 59 -LAAALAAALADHRAVDAFAPVKEVEITQERPTSLIRASAAALVCVGAIGVHFR-- 117
QY 118 -RPERVSTAAALASQCPVAVIYPRHPIPIGRDAWIVYVADSSDGLVLCGAAWAE 175
DB 118 GGP-----AITSIVRSALCPVAVVQAPSLPAARVSGVAEVDN-----GTVLRHAFEEA 169
QY 176 RLADSPRYVTCRQSGVGTGDDVRAASLDRLMARMPRPYDPVRVQSAAVHGLLDYLAG 235
DB 170 RLNG---VGLC---AVGN--PSARVELERRLRMRLTPDVQASAVLVGVEQHLRAD 220
QY 236 GRSVHNVVLSASDQEHVEQIVGAPGNAVL 264
DB 221 HRAGRLV---TDVYRAALCHA--GHSVL 245

RESULT 8
Y2005_MYCTU
ID Y2005_MYCTU STANDARD; PRT; 295 AA.
AC P64921; Q10851.
DT 01-OCT-1996 (Rel. 34, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Hypothetical protein Rv2005c/MT2061.
GN OrderedLocustNames=Rv2005c; MT2061; ORFNames=MYC39.12;
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriineae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium tuberculosis complex.
OX NCBI_TaxID=1773;
RN
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=634230; DOI=10.1038/31159;
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C.M.,
RA Harris D.E., Gordon S.V., Eiglmeyer K., Gae S., Barry C.E. III,
RA Tekala F., Badcock K., Basham D., Brown D., Chillingworth T.,
RA Connor R., Davies R.M., Devlin K., Fellwell T., Gentles S., Hamlin N.,
RA Holroyd S., Hornsby T., Jagsels K., Krogan A., McLean J., Moule S.,
RA Murphy L.D., Oliver S., Osborne K., Quail M.A., Rajandream M.A.,

```

RA Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544(1998).  
RN [2]  
RC NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RP STRAIN=CDC 1551 / Oshkosh;  
RX MEDLINE=22206494; PubMed=12218036;  
DOI=10.1128/JB.184.19.5479-5490.2002;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J.D., DeBoy R.T., Dodson R.J., Gwinn M.L., Haft D.H.,  
RA Hickey E.K., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D.,  
RA Salzberg S.L., Delcher A., Ustcherback T.R., Weidman J.F., Khouri H.M.,  
RA Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,  
RA Fraser C.M.;  
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL J. Bacteriol. 184:5479-5490(2002).  
CC -1- SIMILARITY: To B.subtilis yx1E.  
CC -----  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use as long as its content is in no way modified and this statement is not  
CC removed.  
CC -----  
CC EMBL; BX842578; CA98383.1; -; Genomic DNA.  
DR EMBL; AE000516; AAK46338.1; -; Genomic DNA.  
DR PIR; C70759; C70759.  
DR TIGR; MT2061; -  
DR TubercuList; RV2005c; -  
DR InterPro; IPR006015; USP.  
DR InterPro; IPR006016; USPA.  
DR Pfam; PF00582; USP. 2.  
DR PRINTS; PRO1438; UNVERSLSTRESS.  
DR Complete proteome; Hypothetical protein.  
SQ SEQUENCE 295 AA; 30985 MW; 827D60452E24BE33 CRC64;  
  
Query March 18.2%; Score 257; DB 1; Length 295;  
Best Local Similarity 29.4%; Pred. No. 7.2e-12;  
Matches 91; Conservative 43; Mismatches 126; Indels 50; Gaps 9;  
  
QY 1 MNOSHKPPSIYVIGDGSKPAVQAALMAVDEASRDIPRLLYAIEPD----- 47  
DB 1 MSKRKQHGIVGVGDGSLSDAACMGATDAAMENIPLTVHVNVNADVAITPMPYPETW 60  
QY 48 -----DPGYAAGAAARKLAANAARYAFTAVEADRPVKEVEITQERPVTSILIRASA 102  
DB 61 GVMQEDSG-----RQIVA--NAVKLAKAV--GADRKLSVSKSELVSTPPTWVEISN 109  
QY 103 AALAVCGAIGVHHFRPRVVGSTAAALSLAQCFAIVR-----PHRVP--IGRDAA 152  
DB 110 EAENVVLGSSGRGALARGLSVSSSLVRRAGCFVAVHSDDAVI PPOHAPVVLVGIDGS 169  
QY 153 WI-----VVEADGSSDGVLLGAVMAEARLRDSPRVVTCRQSGVGTGDDVRASLDRL 207  
DB 170 PVSELATAVAFDEASRRGVELIAVHAW-----SVEVEVLELGLDPSAVQDEAEISLBERL 224  
QY 208 ARWQPRYPDVVQSAAHVHGEILDYLAGRSVHVNVLSASDOEHVE-QLVGAFGNAVLYOE 266  
DB 225 AGWQERYPDVPSRVVVCDRPARKLVQKSASQQLVVGSHGRGGLTGMLGVSNAVLHA 284  
QY 267 AGCTLLVVGQ 276  
DB 285 ARVPVIVARQ 294  
  
RESULT 9  
Y2028 MYCBO STANDARD; PRT; 295 AA.  
AC P64922; Q10851;  
DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-MAY-2005 (Rel. 47, Last annotation update)  
DB Hypothetical protein MB2028c.  
GN OrderedlocusNames=MB2028c;  
OS Mycobacterium bovis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;  
OC Mycobacterium tuberculosis complex.  
RX NCBI\_TaxID=1765;  
RN [1]  
RC NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RP STRAIN=AP2122/97;  
RX MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;  
RA Garnier T., Sigmeier K., Camus J.-C., Medina N., Mansoor H.,  
RA Pryor M., Duthey S., Grondin S., Lacroix C., Monsemp C., Simon S.,  
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
RT "The complete genome sequence of Mycobacterium bovis."  
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
CC -1- SIMILARITY: To B.subtilis yx1E.  
CC -----  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use as long as its content is in no way modified and this statement is not  
CC removed.  
CC -----  
CC EMBL; BX248341; CAD96881.1; -; Genomic DNA.  
DR InterPro; IPR006015; USP.  
DR InterPro; IPR006016; USPA.  
DR Pfam; PF00582; USP. 2.  
DR PRINTS; PRO1438; UNVERSLSTRESS.  
DR Complete proteome; Hypothetical protein.  
SQ SEQUENCE 295 AA; 30985 MW; 827D60452E24BE33 CRC64;  
  
Query March 18.2%; Score 257; DB 1; Length 295;  
Best Local Similarity 29.4%; Pred. No. 7.2e-12;  
Matches 91; Conservative 43; Mismatches 126; Indels 50; Gaps 9;  
  
QY 1 MNOSHKPPSIYVIGDGSKPAVQAALMAVDEASRDIPRLLYAIEPD----- 47  
DB 1 MSKRKQHGIVGVGDGSLSDAACMGATDAAMENIPLTVHVNVNADVAITPMPYPETW 60  
QY 48 -----DPGYAAGAAARKLAANAARYAFTAVEADRPVKEVEITQERPVTSILIRASA 102  
DB 61 GVMQEDSG-----RQIVA--NAVKLAKAV--GADRKLSVSKSELVSTPPTWVEISN 109  
QY 103 AALAVCGAIGVHHFRPRVVGSTAAALSLAQCFAIVR-----PHRVP--IGRDAA 152  
DB 110 EAENVVLGSSGRGALARGLSVSSSLVRRAGCFVAVHSDDAVI PPOHAPVVLVGIDGS 169  
QY 153 WI-----VVEADGSSDGVLLGAVMAEARLRDSPRVVTCRQSGVGTGDDVRASLDRL 207  
DB 170 PVSELATAVAFDEASRRGVELIAVHAW-----SVEVEVLELGLDPSAVQDEAEISLBERL 224  
QY 208 ARWQPRYPDVVQSAAHVHGEILDYLAGRSVHVNVLSASDOEHVE-QLVGAFGNAVLYOE 266  
DB 225 AGWQERYPDVPSRVVVCDRPARKLVQKSASQQLVVGSHGRGGLTGMLGVSNAVLHA 284  
QY 267 AGCTLLVVGQ 276  
DB 285 ARVPVIVARQ 294  
  
RESULT 10  
Q7WZ49\_9ACTO PRELIMINARY; PRT; 281 AA.  
AC Q7WZ49;  
DT 01-OCT-2003 (Tremblrel. 25, Created)  
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DB Hypothetical protein.  
OS Nonomuraea sp. ATCC 39727.



```

Db      68  |VGTAAERRRYDGLTVDITVPAEAVSALLAEGRHATVAVTSGRGGELKGLLGSVS 127
Qy      127  |AALASQCPVAIVRPHRPVIGRDAAMIWEADSSDIGVLLGAMAEARLRDSPVRVT 186
Db      128  |LAVASRADCPVVVVRGKESKALSGSHERVLLAGBDPTSGAAVRAFRADVRGCELDYR 187
Qy      187  |C-----RQSGVGDGTDD-----VRASIDRWLARWQPRYPDVVRQSAAVHGEILDYL 232
Db      188  |AMRCPAYENADEGAPSDSEDPERRASALIDTLVAEAAHEHPSVRLKRTIEGPARKYL 247
Qy      233  |AGLRSHVMVVLASDQE-HYEQLVGAPAGNAVLEAGETLLVVGQ 276
Db      248  |VHRTAADLVVVGARHRSGRFGLQIGRVHTLLOHAAACPVAVVQ 292

```

## RESULT 13

```

Q73262_MYCPA PRELIMINARY; PRT; 294 AA.
ID Q73262_MYCPA
AC 073262;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=MAP1741c;
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium avium complex (MAC).
OX NCBI_TaxID=1770;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=K10;
RA Li L., Bannantine J., Zhang Q., Amonsin A., Alt D., Kapur V.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017233; AF04058.1; -; Genomic DNA.
DR GO; GO:0006950; P:response to stress; IEA.
DR InterPro; IPR06015; Usp.
DR InterPro; IPR06016; UspA.
DR Pfam; PF00582; Usp; 2.
DR PRINTS; PRO1438; UNRS1STRESS.
KM Complete proteome.
SQ SEQUENCE 294 AA; 31246 MW; B86F65074298A658 CRC64;

```

Query Match 15.2%; Score 214.5; DB 2; Length 294;  
 Best Local Similarity 28.0%; Pred. No. 1.5e-08;  
 Matches 86; Conservative 47; Mismatches 123; Indels 51; Gaps 12;

```

Qy      1  |MNQSHKPSIVIGIDSKPAVOALMAVDEASRDIPRLLYALE-----PDDPGYAHG 55
Db      1  |MAASKSGCVLVGVDSPPASFAVCMAARDAKENVPLTLVHMVNAATVWFOVE-MAEA 59
Qy      56  |AAARKLAANAENAVRYAFTAVEAD-----RPVKVEVETIOERPYTSLIRASAAALVCV 109
Db      60  |VAMQE-----DGRRVLEQAVXIADATANGRKLAITTELMHAPAPATLAQISEAEVLVV 115
Qy      110  |GAIGVHFRPRVSGSTAALALLSAQCPVAIVR-----PHRVPI-----GRDAWI-- 154
Db      116  |GSTRGAIGRLLLSVSSGLVRRAKCPVAIVIHDEDPMPYRQARVLVGVIDGSPASELAT 175
Qy      155  |VEADSSDIGVLLGAVMAEARLRDSPVRVVTTCRQSGVGDGTDD-----VRASIDRWLAR-- 209
Db      176  |AIADSEASRRGVDLNAVHAMS---DTQV-----FGLRGIDMPVNRSEAEKSLAERL 223
Qy      210  |--WQPRYPDVVRQSAAVHGEILDYLAGRSVHMVVLASDQEHVE-QLVGAPGNVLOE 266
Db      224  |AGMQRERYPDVVHMVVCDRPARQLIEQSESAQLTVVSGHGGIAGTLLGSVENAVVHS 283
Qy      267  |AGCTLLV 273
Db      284  |VRMPVIV 290

```

```

RESULT 14
Q77Y75_MYCBO PRELIMINARY; PRT; 297 AA.
ID Q77Y75_MYCBO
AC 077Y75;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Hypothetical protein TB31.7.
GN Name=TB31.7; OrderedLocustNames=Mb2656;
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium tuberculosis complex.
OX NCBI_TaxID=1765;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=AF2122/97;
RC MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;
RA Garnier T., Eigimeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthey S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
DR EMBL; BX248343; CAD94841.1; -; Genomic DNA.
DR GO; GO:0006950; P:response to stress; IEA.
DR InterPro; IPR06015; Usp.
DR InterPro; IPR06016; UspA.
DR Pfam; PF00582; Usp; 2.
DR PRINTS; PRO1438; UNRS1STRESS.
KM Complete proteome.
SQ SEQUENCE 297 AA; 31652 MW; A590F7058D1E8695 CRC64;

```

Query Match 15.2%; Score 214; DB 2; Length 297;  
 Best Local Similarity 27.2%; Pred. No. 1.6e-08;  
 Matches 84; Conservative 37; Mismatches 134; Indels 54; Gaps 10;

```

Qy      1  |MNQSHKPSIVIGIDSKPAVOALMAVDEASRDIPRLLYALEP-----DPGY 51
Db      1  |MSGNSSLGIIVIGIDSPBAQVAVRMAARDAELRKIPTLVHVASPEVATWLEVPPLPGV 60
Qy      52  |AA---HGAAARKLAANAENAVRYAFTAVEAD---RPVKVEVETIOERPYTSLIRASAAA 104
Db      61  |LRMQQDHG-----RHLIDDAIKVVEQASLRAGPPTVSHSIVPAAAPTLVDMSKDA 111
Qy      105  |ALVGVGAIQVHFRPRVSGSTAALALLSAQCPVAIV-----RPHRVYIGRDAIVTV 156
Db      112  |VLMVVGCLSGRWPFGRLLSSVSSGLRHAHCAPVITIHDEDSVMPHPQAP-----VLV 164
Qy      157  |EADSSDIGVLLGAVMAEARLRDSPVRVVTTCRQSGVGD-----GDDVFRAS-----LDR 205
Db      165  |GVDSASASELATATAPPEASRRN--VDLVALHMSDVTSEWFGIDMPAQSMAEQVLAE 222
Qy      206  |WLRWQPRYPDVVRQSAAVHGEILDYLAGRSVHMVVLASDQ-EHVEQLVGAPGNVLA 264
Db      223  |RLAGMQRERYPVNAITRVVADQPARQLVGRSEBAQLVVGSRGRGYAGLTVSGERTVA 282
Qy      265  |QEAGCTLLV 273
Db      283  |QLARTPVIV 291

```

## RESULT 15

```

Q06189_MYCTU PRELIMINARY; PRT; 297 AA.
ID Q06189_MYCTU
AC 006189; Q7D6V7;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 13-SEP-2005 (TReMBLrel. 31, Last annotation update)
DE Hypothetical protein TB31.7 (Universal stress protein family).
GN Name=TB31.7; OrderedLocustNames=MT2698, RV2623;
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinomycetales;

```





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using SW model

Run on: March 23, 2006, 05:24:53 ; Search time 19.0287 Seconds  
(without alignments)  
1212.194 Million cell updates/sec

Title: US-10-617-038-21  
Perfect score: 1410  
Sequence: 1 MNQSHKPSIVVIGIDSKPA.....GNAVLOAGCTLLVVGQOYL 279

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/1/1aa/5\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/6\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/H\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/PCNUS\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/RE\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfiletest.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	207	14.7	301	2	US-09-605-703B-1870
2	115	8.2	3562	2	US-09-679-279-14
3	107	7.6	815	2	US-09-252-991A-27251
4	106.5	7.6	3546	2	US-09-679-279-13
5	105.5	7.5	426	2	US-09-489-039A-9821
6	102	7.2	458	2	US-09-252-991A-28857
7	102	7.2	1650	2	US-09-252-991A-21798
8	100	7.1	1277	2	US-09-602-787A-54
9	100	7.1	2012	2	US-09-602-787A-50
10	99.5	7.1	1927	2	US-10-152-886-63
11	99	7.0	1138	2	US-09-252-991A-20291
12	99	7.0	2162	2	US-09-477-962-97
13	98.5	7.0	381	2	US-09-252-991A-26281
14	98	7.0	323	2	US-09-252-991A-17111
15	97	6.9	348	2	US-09-489-039A-10294
16	97	6.9	750	2	US-09-252-991A-32334
17	97	6.9	736	2	US-09-252-991A-32256
18	96.5	6.8	849	2	US-09-252-991A-21865
19	96.5	6.8	1996	1	US-08-804-227C-9
20	96.5	6.8	1996	1	US-08-804-198-3
21	96.5	6.8	3491	1	US-07-642-734C-2
22	96.5	6.8	3491	1	US-08-439-009A-2
23	96	6.8	433	2	US-09-351-150A-19
24	95.5	6.8	280	1	US-09-328-352-4875
25	95.5	6.8	3170	1	US-07-642-734C-5
26	95.5	6.8	3170	2	US-08-439-009A-5
27	95.5	6.8	4472	1	US-08-804-227C-2

28	95	6.7	1721	2	US-10-042-665A-6	Sequence 6, Appl
29	94.5	6.7	655	2	US-09-252-991A-31645	Sequence 31645, A
30	94.5	6.7	659	2	US-09-902-540-16383	Sequence 16383, A
31	94.5	6.7	735	2	US-09-252-991A-30569	Sequence 30569, A
32	94	6.7	367	2	US-09-902-540-15067	Sequence 15067, A
33	94	6.7	613	2	US-09-758-759-151	Sequence 151, App
34	94	6.7	628	2	US-09-252-991A-29296	Sequence 29296, A
35	94	6.7	1271	1	US-08-095-734-2	Sequence 2, Appl1
36	94	6.7	1271	1	US-08-444-623-2	Sequence 2, Appl1
37	94	6.7	1271	2	US-08-471-869-2	Sequence 2, Appl1
38	94	6.7	1271	2	US-09-342-563-2	Sequence 2, Appl1
39	94	6.7	1271	2	PCT-US94-08267-2	Sequence 2, Appl1
40	93.5	6.6	293	2	US-09-252-991A-22215	Sequence 22215, A
41	93.5	6.6	426	2	US-09-902-540-11847	Sequence 11847, A
42	93.5	6.6	569	2	US-09-252-991A-20194	Sequence 20194, A
43	93	6.6	407	2	US-09-252-991A-19349	Sequence 19349, A
44	93	6.6	3567	1	US-07-642-734C-4	Sequence 4, Appl1
45	93	6.6	3567	2	US-08-439-009A-4	Sequence 4, Appl1

## ALIGNMENTS

```
RESULT 1
US-09-605-703B-1870
; Sequence 1870, Application US/09605703B
; Patent No. 6962989
; GENERAL INFORMATION:
; APPLICANT: Pompeius, Markus
; APPLICANT: Krieger, Burkhard
; APPLICANT: Schröder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauser, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: BGI-1292C
; CURRENT APPLICATION NUMBER: US/09/605, 703B
; CURRENT FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: 60/142,764
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: 60/152,318
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 2934
; SEQ ID NO 1870
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-605-703B-1870
Query Match 14.7%; Score 207; DB 2; Length 301;
Best Local Similarity 26.1%; Pred. No. 8.2e-14;
Matches 77; Conservative 46; Mismatches 14; Indels 38; Gaps 10;
QY 10 IVVIGIDSKPAVQALMAVDEAASRDIPLRN-----LYA---IEPDDGYANGAA 57
DB 7 VVVAVDSDDSKQAVRWNAANTANKRGIPLRASSTYTPQGFYAGMWPPDELFDLQAEA 66
QY 58 ARKLAALBNVRYAFTVAEADRPVYKEVELTORPPTSILIRAAALVCGAIGVHHF 117
DB 67 LEKLINEARD-----IAHEVAPE- IKIGHTAESSPTDMLLEMSPDATMTVMGSRGAGGL 119
QY 118 RPERVSTAAALALASQCPVAIVRPHRVPIGRDAAM--IIVADSSSDIGVILGAVAAEA 175
DB 120 SGWMSVSGAVSHACPPVYVAREDA- VNEBSKTEPVVGVVDGSEVSQATETAPAEA 178
QY 176 RLR-----DSPRVVTCRQSGVGDTGDDV-RASLDRLIARQP--RYPDYRVQ 220
DB 179 EARGAEIVAVHTWMDQVQSLAGIAAQQOMDEVERQQTDMILIERLAPVEKYPSTYVK 238
QY 221 SAAHGELDLVTLAIGRSYHHVTLASADQEHVE-QLVGACGNVLOAGCTLLVV 274
DB 239 KITRDRPVRALAEASBNADLLVVGSHRGQFGKMLIGSTSRALLQSPCPMMVV 293
```

```
RESULT 2
US-09-679-279-14
; Sequence 14, Application US/09679279
; Patent No. 6524841
; GENERAL INFORMATION:
; APPLICANT: McDaniel, Robert
; TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic
; FILE REFERENCE: 30622004700
; CURRENT APPLICATION NUMBER: US/09/679,279
; CURRENT FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/158,305
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 60/190,024
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 3562
; TYPE: PRT
; ORGANISM: Micromonospora megalomicea
US-09-679-279-14

Query Match      8.2%; Score 115; DB 2; Length 3562;
Best Local Similarity 27.4%; Pred. No. 0.023;
Matches 95; Conservative 25; Mismatches 119; Indels 108; Gaps 20;

QY 13 GIDSKRAVQALMAVDEAASRDILRLLYA-----IERDDGYAAH-GA----- 56
D 2743 GDDVPRPGA-AAVGVLCRAQAESPDREVLVDGPETPRPAVDPQLVRDGAFAVFRILT 2801
QY 57 --AAKLAAMENAVR-----YAFTAVEADRP-----VKVEVEITQER----- 92
D 2802 PLAGPVAVADRAVRLVPGNGSGIEAVAFAVVPADRPPLAEVRAVRAIGVFRDYL 2861
QY 93 -----PYTSLIRASAAALVCVGAIGVHHFRPERVGTAAALASAOCPVAIYRPHR--- 144
D 2862 ALGVYPRPAEWGTASGVTEVGS-GVYRFTP---GQAVTGLFGAGRGPAV-VADHRLLT 2916
QY 145 -VPIGRDAWTVVNA-----DSSSDI-----GVTLGAVM----- 172
D 2917 PVPBG-----WAVVAAVPIAFTTAHYALHDLAGLQGSVLVHAAAGVGMAAVLARR 2972
QY 173 --AEARLRDSFVRVVTGROSGVGTGDDVRASL-----DRWLAWQPRYPDVVQSAVH 225
D 2973 AGAEVPAATASPAKPIRLAIGLDD--DHIASSRESGGERPAARTGKGVVDVLNS--LT 3028
QY 226 GELID---YLAGIGRSVHVVLTSASDQEHVEQLVGAPGNAVLOEAG 268
D 3029 GDILDESARLADGDGVFVEM---GKTDLRPAEQFRGRYVPEDLAEAG 3072

RESULT 3
US-09-252-991A-27251
; Sequence 27251, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27251
; LENGTH: 815
; TYPE: PRT
```

```
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27251

Query Match      7.6%; Score 107; DB 2; Length 815;
Best Local Similarity 26.9%; Pred. No. 0.019;
Matches 87; Conservative 31; Mismatches 107; Indels 98; Gaps 19;

QY 4 SHKPSIVGIDDSKPA-----VQALMAVDEAASDILRL-----LVAIERDDP 49
D 293 SLSEPLAAMAAALAGLAPRGGRWRRLAQALLAVLGCAVVLHRLRORPLDPAARAGP 352
QY 50 GY-----AAGGAARKLALEN-----AVRYAFTAVEADRPVKVEVEITQERPYT 95
D 353 GVARRRPAARAAAGRLAAAWRVGLGETDPIAAR-AVPRPLAAGLARAEV---HRRAG 408
QY 96 SLIRASAAALVCVGAIGVHHFRPERVGT-----AAALAS--AOCPVAIYR----- 141
D 409 QRRRLAARAAAGVLAEEAVLGSVLQRRRAGATGGQGRRLVASRRPAGDPLAFVRLARRG 468
QY 142 -----PHRYPI--GRDAWTVVEADSSDILGLAVMAEARLRDSFVRVVTGROS 150
D 469 APALALPAAAGTVPAMPGRGAAMLRL---PREVGAAG--TRADARVVAAGVRA-----A 518
QY 191 GVGDGDDVRAASLDRLAWQPRYPDVVQSAAV-HG-----ELLDY---LAGLG 236
D 519 GI-----DGAQAGL-----AVEVQAVVGHAGSEDLALVVEVLDAHLFQSLG 562
QY 237 RSV-HMVVLSASDQEHVEQLVGA 258
D 563 DSLGRLLALELVDSHADQVDA 585

RESULT 4
US-09-679-279-13
; Sequence 13, Application US/09679279
; Patent No. 6524841
; GENERAL INFORMATION:
; APPLICANT: McDaniel, Robert
; TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic
; FILE REFERENCE: 30622004700
; CURRENT APPLICATION NUMBER: US/09/679,279
; CURRENT FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/158,305
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 60/190,024
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 3546
; TYPE: PRT
; ORGANISM: Micromonospora megalomicea
US-09-679-279-13

Query Match      7.6%; Score 106.5; DB 2; Length 3546;
Best Local Similarity 23.9%; Pred. No. 0.18;
Matches 81; Conservative 33; Mismatches 120; Indels 105; Gaps 18;

QY 19 PAVQALMAVDEAASR-----DILRLIYAE--PDPRGYAAHGAARKLAA-----AENAV 68
D 1425 PFGREFFWLEPNPARVADSDVSLRYRLEWHPTPGEGRLDGTWLLATYFGRADRV 1484
QY 69 RYAFTAVEADRPVKVEVEITQERP-----VTSIRASAAALVCVGAIGVHHFRPERVG 123
D 1485 EAARQALIESAG--ARVEDLVVERTRGVRLVRLDVGVPVAGVLCLEPAAH-----EPAAEH 1538
QY 124 STAAALAS-----AOCPVAIYRPHRPIG-----RD-----AAW-----IVV 156
D 1539 SPLAVTSLSDTLDTLQAVAGSGRECPIMVYTEVAVAVGPERLRDPAHGALMALGRVAL 1598
QY 157 E-----ADGSSDIGVLL-GAVMAEARLR-DSFVRVVTGROSGVGTGDD 198
```

Db 1599 ENPAWGGVLDPVSGVAELSRHLGTTLSGAGEEDVALRPGTYARWCR-AGAGGTG-- 1655  
Qy 199 VRASLDNR-----LARMQRPYDVVQSAAYHG-----ELLDYLA 233  
Db 1656 -----RQPRGTUVTYGTGCGVGNHVARMLARQSTPCLVTLASRRGPADGVEELITELA 1709  
Qy 234 GLGRSVHNVVLASDOEHVEQLVGA-----PGNAVLOEA 267  
Db 1710 DLGTRATVACDVTDREQRLALLATVDEHPLSAVFHVA 1748

RESULT 5  
US-09-489-039A-9821

; Sequence 9821, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 9821  
; LENGTH: 426  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-9821

Query Match 7.5%; Score 105.5; DB 2; Length 426;  
Best Local Similarity 23.6%; Pred. No. 0.011; Indels 41; Gaps 8;  
Matches 53; Conservative 38; Mismatches 93;

Qy 57 AARKLAENAVRYAFTVAEADRPVKEVEITQRPVTLIRASAAALVCGAIGVH 116  
Db 36 AGVSLAEAGMGEVAFVTAQPS--TARALINSDBRAMTAAADKVALLRGATFVPH 93  
Qy 117 FRPERVGT-----AAALASQCPVALYRPHRVIGDAMIVVADGSSDIGVL 168  
Db 94 WQVKKISTLTGNPGCELEAMMAAGCRMVAVAPAYPAAGHRT-----RDGRCVH--- 144  
Qy 169 GAWAAEARLRDSPRVVTCROSGVGDGDDVRASLDRLARMPRPYDVVQSAAYHGEL 228  
Db 145 GVPIDQTEFPASDPKTPVS-----PAETSETIA-MQSKLPCLTINA---GQL 186  
Qy 229 LDVLAGLGRSVHNVVLASDOEHVEQLVGAPEGNAVLOEAGCTLLV 273  
Db 187 PALATAGEEKRVLLVDWMEDSHLDQV-----DANAPHARETLIV 227

## RESULT 6

US-09-252-991A-28897  
; Sequence 28897, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 28897  
; LENGTH: 458  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-28897  
Query Match 7.2%; Score 102; DB 2; Length 458;  
Best Local Similarity 22.6%; Pred. No. 0.029;  
Matches 75; Conservative 42; Mismatches 117; Indels 98; Gaps 16;

Qy 27 AVDEASRDIPL--RLIYAIPEDDPGYAAAGAAARKLAENAVRYAFTVAEADRPV 84  
Db 26 AATSSATRAFALATSLIYACAPASWTASQSPQVWMEDFLPENVPALHLDQVTRQ 85  
Qy 85 EVELT-----QRPV---TSLIRASAAALVCGAIGVHRRPRV----- 122  
Db 86 LARISLGTRVRYLVANTYGRPRIRGRATLAKPGRSAVAATGSLDVTGCRRTATIA 145  
Qy 123 -GST-----AAALASQCPVALYRPHRVPI-----GRDAWIVV-----E 157  
Db 146 PGATLLDPPVAMPFPAALSQVLSLYLPAATFMEFTHDGTGTGTVIAGEQTSALQDA 205  
Qy 158 ADG-SSDIGVLAGVAAEARLRDSPRVVTCROSGVGT--GDVVASLDRLARMPRPY 215  
Db 206 ADSQTSARILLTGILVET--ETATRTLVY---LQDSITDGASASLDR-----NSRWP 253  
Qy 216 DVYVQSAAYHG-----ELL-----DYLAGLGRSVHNVVLASDOE 250  
Db 254 DFLAERLAPHGVAVAVNAGISGARLLSDGMSALALARDVLAQPGASSWVMGLIND-- 311  
Qy 251 HVEQLVGAPEGNAVLOEAG---CTLLVVGQOYL 279  
Db 312 -----TAMPGTAFARNNAAPPTLEALLINGYROL 338

## RESULT 7

US-09-252-991A-21798  
; Sequence 21798, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21798  
; LENGTH: 1650  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21798

Query Match 7.2%; Score 102; DB 2; Length 1650;  
Best Local Similarity 23.5%; Pred. No. 0.18;  
Matches 87; Conservative 28; Mismatches 114; Indels 142; Gaps 19;

Qy 10 IVVGID-GSKPAGVQALMAVDEASRDIPRL-----LYAIPEDD----- 48  
Db 1184 IATGLDHGAVPALQOAVLAVQARLDPOFALSEQLAATVVOVSADVDPDRPLQAAALV 1243  
Qy 49 -----PGVAHGAARAKLAENAVRYA-FTVAEAA--DRPVKEVEITQ----- 91  
Db 1244 QRLARLAFQRAAGHQAAGVAGAAAGVAAAGVAPLVRDLNAGPHPEVNEARGVGRAPFA 1303  
Qy 92 ----RPVTLIRASAAALVCGV-AIGVHNF-RPERV-----GSTAAALASQCPVALIV- 140  
Db 1304 RRGALPYDQFLRAADLGAATGVGLAVQTHGIRGLGVALLGAGLGAQAGVADADVAAG 1363  
Qy 141 -----RPHRVPIGRDA-----WIVVADGSS----- 162  
Db 1364 GTVAGQPHPAL--LGADQADRPGVHPAQRRAVDRQLMLAALIGARGQLQAVGLDVVAAG 1422

QY 163 -----DIGVLGAVMAEARLDSFVRVYTCQSGVGTGDDVNASLDRMLARQPRY 214  
Db 1423 DDEVEALVEIGVEIGAAAGDVELEFE-----VADVQGAFFD--DDV-----1460  
QY 215 PDVRVOSAAYHGEILDVLA-----GLGRSVHMYVLSASD-----OEHEVOL-----VG 257  
Db 1461 -----AAVHLEALQLALVADRILAGEGARARVDEAAALAGDAVNRVGDHPRRLPHGLG 1513  
QY 258 APGNAYLQBAQ 268  
Db 1514 VAGQGAAYVAG 1524  
RESULT 8  
US-09-602-787A-54  
Sequence 54, Application US/09602787A  
Patent No. 6696561  
GENERAL INFORMATION:  
APPLICANT: Pompejus, Mark  
APPLICANT: Krüger, Burkhard  
APPLICANT: Schöder, Hartwig  
APPLICANT: Zelder, Oskar  
APPLICANT: Haberbauer, Gregor  
TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS  
TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE  
FILE REFERENCE: BGI-125CP  
CURRENT APPLICATION NUMBER: US/09/602, 787A  
PRIOR FILING DATE: 1999-06-23  
PRIOR APPLICATION NUMBER: USSN 60/141031  
PRIOR FILING DATE: 1999-06-25  
PRIOR APPLICATION NUMBER: DE 19931454.3  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931478.0  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931563.9  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19932122.1  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932124.8  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932125.6  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932128.0  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932180.9  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932182.5  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932190.6  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932191.4  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932209.0  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932212.0  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932227.9  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932228.7  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932229.5  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932230.9  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 1993227.3  
PRIOR FILING DATE: 1999-07-14  
PRIOR APPLICATION NUMBER: DE 19933005.0  
PRIOR FILING DATE: 1999-07-14  
PRIOR APPLICATION NUMBER: DE 19933006.9  
PRIOR FILING DATE: 1999-07-14  
PRIOR APPLICATION NUMBER: DE 19940764.9  
PRIOR FILING DATE: 1999-08-27

PRIOR APPLICATION NUMBER: DE 19940765.7  
PRIOR FILING DATE: 1999-08-27  
PRIOR APPLICATION NUMBER: DE 19940766.5  
PRIOR FILING DATE: 1999-08-27  
PRIOR APPLICATION NUMBER: DE 19940830.0  
PRIOR FILING DATE: 1999-08-27  
PRIOR APPLICATION NUMBER: DE 19940831.9  
PRIOR FILING DATE: 1999-08-27  
PRIOR APPLICATION NUMBER: DE 19940832.7  
PRIOR FILING DATE: 1999-08-27  
PRIOR APPLICATION NUMBER: DE 19940833.5  
PRIOR FILING DATE: 1999-08-27  
PRIOR APPLICATION NUMBER: DE 19941378.9  
PRIOR FILING DATE: 1999-08-31  
PRIOR APPLICATION NUMBER: DE 19941379.7  
PRIOR FILING DATE: 1999-08-31  
PRIOR APPLICATION NUMBER: DE 19941395.9  
PRIOR FILING DATE: 1999-08-31  
PRIOR APPLICATION NUMBER: DE 19942077.7  
PRIOR FILING DATE: 1999-09-03  
PRIOR APPLICATION NUMBER: DE 19942078.5  
PRIOR FILING DATE: 1999-09-03  
PRIOR APPLICATION NUMBER: DE 19942079.3  
PRIOR FILING DATE: 1999-09-03  
PRIOR APPLICATION NUMBER: DE 19942088.2  
PRIOR FILING DATE: 1999-09-03  
NUMBER OF SEQ ID NOS: 678  
SEQ ID NO 54  
LENGTH: 1277  
TYPE: PRF  
ORGANISM: Corynebacterium glutamicum  
US-09-602-787A-54  
Query Match 7.1%; Score 100; DB 2; Length 1277;  
Best local Similarity 21.6%; Pred. No. 0.21;  
Matches 55; Conservative 47; Mismatches 113; Indels 40; Gaps 9;  
QY 4 SHKEPSIVGIGDSKAVQALMA-----VDEAASDIFLRLLYAIERPDGPAAGAAA 58  
Db 313 AHTAPVPLVG--RAWAPVPAVAVSAPIPTDSASVEGMLSLV-----HLEHHIVL 361  
QY 59 RKLAAMENVRYAFTVEADRP-----VKVEVETQERPVTSIL-----RASAAA 105  
Db 362 KSDVPTDGAUKYSATIDEVVDLGLVYRAEIADEGNLTATLERRAIRRKKNAYA 421  
QY 106 LVYCGAIGVHFRPERVGSAAALASAOCPVAYRPHVPIG-RDA-----NIVYE 157  
Db 422 RTNTSALPTTVDPREBARAVATVVAPESSMRPFAVISGDRNPFIHVSVAASLAGLPGVIVH 481  
QY 158 ADGSSDITGVLL-GAVMAEARLDSFVRVYTCQSGVGT--GDDVNASLDRMLARQPRY 214  
Db 482 GMMTSAIGELIAGAAAFNDSDIQTPAKVVEYTAIYMLAPVLPGEIEFVSERSAVDNRPM 541  
QY 215 PDVRVOSAAYHGEIL 229  
Db 542 GEVRYTATVNGMLV 556  
RESULT 9  
US-09-602-787A-50  
Sequence 50, Application US/09602787A  
Patent No. 6696561  
GENERAL INFORMATION:  
APPLICANT: Pompejus, Mark  
APPLICANT: Krüger, Burkhard  
APPLICANT: Schöder, Hartwig  
APPLICANT: Zelder, Oskar  
APPLICANT: Haberbauer, Gregor  
TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS  
TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE  
FILE REFERENCE: BGI-125CP  
CURRENT APPLICATION NUMBER: US/09/602, 787A

```
; CURRENT FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US9N 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: DE 19931454.3
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931478.0
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931563.9
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932122.1
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932124.8
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932180.9
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932182.5
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932190.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932191.4
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932209.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932212.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932227.9
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932228.7
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932229.5
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932230.9
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932927.3
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19933005.0
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19933006.9
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19940764.9
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940765.7
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940766.5
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940830.0
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940831.9
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940832.7
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940833.5
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19941378.9
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19941379.7
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19941395.9
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19942077.7
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: DE 19942078.5
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: DE 19942079.3
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: DE 19942088.2
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 678
; SEQ ID NO 50

; LENGTH: 2012
; TYPE: PR
; ORGANISM: Corynebacterium glutamicum
US-09-602-787A-50

Query Match
Best Local Similarity 21.6%; Pred. No. 0.4;
Matches 55; Conservative 47; Mismatches 113; Indels 40; Gaps 9;

QY 4 SHKPSIVGIDGSKPAVOALNA-----VDEASRDIPRLTYAIBPDDPGYAAGAAA 58
D 995 AHTAPDVVG--RAMPVFAVKSAVTPTGDSASVEGMTSLV-----HLEHIVL 1043
QY 59 RKLAAENAVRYAFTAVEADRP-----VKVEITQEBPVTSLI-----RASAAA 105
D 1044 KSDVPTGALKVSTADVEDVDTGLRLVIRAEIADAEGLIATLAEFPAIRGKGNVA 1103
QY 106 LVCGAIGVHHFRPVRGSTRALALASQCPVALVRHVRPIG-RDAA-----WAVE 157
D 1104 RTNTSALPTTVDFRSARAATVAVAPESMRPFAVISGDRPHIVSDVAASLAGLPGVIVH 1163
QY 158 ADGSSDVGVL--GAVMAEARLRDSPVRVYTCROSGVGD--GDDVRASLDRLARMPRY 214
D 1164 GMMTSALGELIAGAFNDEQITPAKAVETRTATMLAPVLGEIEFSVERSAVDNRPGM 1223
QY 215 PDVRVQSAVHGELL 229
D 1224 GEVRTVATVNGNLV 1238

RESULT 10
US-10-152-886-63
; Sequence 63, Application US/10152886
; Patent No. 6912470
; GENERAL INFORMATION:
; APPLICANT: ECOPIA BIOSCIENCES INC.
; APPLICANT: Farnet, Chris
; APPLICANT: Staiff, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF ENEDIYNE RING
; FILE OF INVENTION: STRUCTURES
; CURRENT APPLICATION NUMBER: US/10/152, 886
; CURRENT FILING DATE: 2002-05-21
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 63
; LENGTH: 1927
; TYPE: PR
; ORGANISM: Micromonospora megalomicea
US-10-152-886-63

Query Match
Best Local Similarity 24.2%; Pred. No. 0.43;
Matches 67; Conservative 40; Mismatches 85; Indels 85; Gaps 13;

QY 23 AALMAVDEASRDIPRLTYAIBP--DOPGYAAHGAARKLAAENAVRYAFTAVEADRP 81
D 1213 AAMATDSGAS---LALLGRSAPAEDEPLAAN--LSRMTAGVTYRABADVTDPQV 1265
QY 82 VKVEITQEBPVTSLI-----AAALVCGAIGVHHFRPVRGSTRALALASQOC 135
D 1266 RRAVAELITADLGPRTATVILHGAGRERPAALNLDGAVR-RTFAFKLDGLST----- 1316
QY 136 PVAIVRPHRVPIGRDAAMIVEADGSSDIGVLGAVMAEARLRDSPVRVYTCROSGVGD 195
D 1317 -LAAVDDRL-----RLVVTIGSITGRAGLR----- 1341
QY 196 GDDVRASLDRLARMPRYPDVRVQSAVHGE-----LQDY--LAGTGRSVHMYLS 245
D 1342 GEAHYATANDMLA-----EATTAFRHRRHCRSICLEWSVMSGVGKGRLSVE 1390
QY 246 ASDQEHVEQLVGAPGNAVLQ-----EAGCTLLVVGQ 276
```

Db 1391 SLHREGVAALTPDOGVAVLRLRLADPEATGVVWSGR 1427

## RESULT 11

US-09-252-991A-20291  
Sequence 20291, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252, 991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 20291  
LENGTH: 1138  
TYPE: PR  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20291

Query Match 7.0%; Score 99; DB 2; Length 1138;  
Best Local Similarity 23.9%; Pred. No. 0.23;  
Matches 70; Conservative 26; Mismatches 91; Indels 106; Gaps 15;

QY 44 IEPDDPGY---AAHGAARKLAAENAVRYA-----FTAVEADRPVKEVEITQE 91  
DB 562 VEPDRTGATPFGIGIAIVLAADVRFEGAGRADVQLAAGACGCP-----RQ 613  
QY 92 RP-----VTSIRASAAAALVCV-----AIGVHPRP-----RVGSTRALAL 131  
DB 614 RPSRDRVTTPACIGGAFQADVPADLANQOALGAHLLVEEVRQGLDGLLAPLAV 673  
QY 132 SAQCPVAIVRPHRPIGRDAAMIVEADGSSDVLGAVAAEA-----R 176  
DB 674 QVRLGLAV--QHR-----GVVGAVALAADHRYEAGEDEVGSQLH 711  
QY 177 LRDSRVVTCRQSGVDGTGDVRA-----LDRML-----ARKOPRPDVAVQSAV 224  
DB 712 RRVGVPAVDFAAPRGAQVQCGEAGAEHADIFRVLDEAVAGVAGGRRRGLGQCCHVDGEA 771  
QY 225 HGEILDVLAGRGVNHVVSASDQ-----HV-----EGLVG--APGNAV 264  
DB 772 GGEVVDRRHAGRG---ALAAEDRQAAVVHVVGAEIVQFGEVAAQGEAVL 820

## RESULT 12

US-09-477-962-97  
Sequence 97, Application US/09477962  
Patent No. 6927286  
GENERAL INFORMATION:  
APPLICANT: SHEN, BEN  
APPLICANT: DU, LIANGCHENG  
APPLICANT: SANCHEZ, CESAR  
APPLICANT: CHEN, WEI  
APPLICANT: EDWARDS, DANIEL J.  
TITLE OF INVENTION: BLEBOMYCIN GENE CLUSTER COMPONENTS AND THEIR USES  
FILE REFERENCE: 407T-895820US  
CURRENT APPLICATION NUMBER: US/09/477, 962  
CURRENT FILING DATE: 2000-01-05  
PRIOR APPLICATION NUMBER: 60/115,435  
PRIOR FILING DATE: 1999-01-06  
PRIOR APPLICATION NUMBER: 60/118,848  
PRIOR FILING DATE: 1999-02-05  
NUMBER OF SEQ ID NOS: 133  
SOFTWARE: PatentIn Ver. 3.0  
SEQ ID NO 97  
LENGTH: 2162

TYPE: PR  
ORGANISM: Streptomyces verticillius  
FEATURE:  
OTHER INFORMATION: ORF26  
US-09-477-962-97

Query Match 7.0%; Score 99; DB 2; Length 2162;  
Best Local Similarity 24.5%; Pred. No. 0.57;  
Matches 79; Conservative 35; Mismatches 122; Indels 86; Gaps 16;

QY 7 PPSIVGIDG-----SKPAV-----QALMAVDEAASDIPRLIYAIEPDDPGYAAH 54  
DB 256 PABLVAIGDVGAKLSTDPATVTLAAFRIVMRL--NGERMLPVAL-----TRD 302  
QY 55 GAAARKLAAENAVRYAFTAVEADRPVKEVEITQRPVTSIRASAAALVCVAGIV 114  
DB 303 GRSHPELR-----TAGAFERLPLVHEIRHETAFAEYAR--ALDALVAEGEELL 350  
QY 115 HFRPERVGSSTAALALSAQCPVAIVRPHR--VPGR-DAAWIVEADGSSDVLGAV 171  
DB 351 DHCDPELIGS-----LQTAGPCPCTFTNHQAEPPVRAGITFTTVHQDSGTPIPVRLTAR 406  
QY 172 MAEARLR-----DSPVRVVTCRQSGV-----GDTGPDVRAASLRW--LAR 209  
DB 407 RDGARLRMEIGYDEGRIDETFPENAACTRIIEGVVASABEGVQ--DIRMLSDETARLR 465  
QY 210 WQRPYDPVAVQSAVHGEILDVLAGRGSVNHVVSASDQ-----EHVEQ 254  
DB 466 EAGIGRVELPGNAVHLEFPAQNA---RTPGNAVASAGBDALVVALEDRSNRLAHHLTG 522  
QY 255 LVGAPGNALVQENAGCTL-LVVG 275  
DB 523 LGVTGPRHVVVSVGRSHELLVG 544

## RESULT 13

US-09-252-991A-26281  
Sequence 26281, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252, 991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 26281  
LENGTH: 381  
TYPE: PR  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26281

Query Match 7.0%; Score 98.5; DB 2; Length 381;  
Best Local Similarity 26.1%; Pred. No. 0.053;  
Matches 82; Conservative 29; Mismatches 122; Indels 81; Gaps 17;

QY 9 SIIVGIDGSKPA--VOALMAVDEAAS-RDIPRLIYAIEPDDPGYA----- 52  
DB 94 ALVVGFGRPARPAGVNAHPDADGATGADVPEQLAGAGQARQAGADVALQALVAI 153  
QY 53 -----ANGAARKLAAENAVRYAFTAVEADRPVKEV-VEITQRPVTSIR 100  
DB 154 GFVGQARGORAHGAGH--GAAFH--RHQVETGARQORVAGEAVETALERPDAVFOHA 208  
QY 101 SAALAVCVGALGVNHPRPVGSTAAALALSA---QCP-----VAIVRPHRVIGRDAW 153  
DB 209 QATGA-----EATGV-----QQAQDGGRAVAVVVAEDQAPAVVULCICORRQAHQKQKD 259



QY	154	IV---VEADSSDIDGLTGA	WMAEATL	RDSPRVVT	CRQSGVGDTDD	DVR-----	ASL	203	
Db	260	LLQGFAEAGGGGREG-----	GGRRDP	QFG---R	ROLPESTADAE	QHREVAAGQHA		307	
QY	204	DRWLRLKWDPRYPDVR	QGSAAVHG	ELDYL	AGRSVHM	VYLSASDQEH	EQLGAPGN	263	
Db	308	GRLLAPGQWQF	GGERRRR	PFAGS----	ADARRQ	QFOLARRR	DDPAGLQ--	GTGP--L	357
QY	264	LOEAGCTLLV	VGQ	277					
Db	358	LAEGITVL	ADAHQ	371					

```

RESULT 14
US-09-252-991A-1711
; Sequence 1711, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIORITY FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIORITY FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIORITY FILING DATE: 1998-07-27
; NUMBER OF SEQ. ID NOS: 33142
; SEQ ID NO 1711
; LENGTH: 323
; TYPE: PRF
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-1711

```

		7.0%;	Score 98;	DB 2;	Length 323;
		Best Local Similarity 27.7%;	Pred. No. 0.047;		
		Matches 61;	Conservative 22;	Mismatches 81;	Indels 56;
					Gaps 10;
Qy	29	DEAASRDIPRLYLAIKP----	DDPGYAHGAARKLIA-AAENAVRYATFVAEADRPV	82	
Db	86	DIAEATQDPKVFQDGIHPLRRVGRMAGVAHIAAGQLATVAEHNLQGLADDAQCRILV	145		
Qy	83	KVEVEIDPERVPTSLIRASAAAALVCYGAIGHV-----	HPRE	120	
Db	146	RTPIQRREQR-----	TDQAADPLVYGERQVHROAQRRGERDRGRHQAQVALHVR--	197	
Qy	121	RVGSTAAALALS-----	AACPVALVHPHVPV-I-GRDAMIIVYEDAGSSDIGLLGAVMAE-	174	
Db	198	RATVEVAAILALAEERRRHPAIGLAVRRHHVGMPEKADATVAGADGIRIG--	LAAGVEE	255	
Qy	175	-----	ARLRDSPRVVTCROSGVG-----	DTGDDVVAS	202
Db	256	QFAVDAETROVADVADROQVGLAADGVGEHQVGGVADVAAP	295		

RESULT 15  
 US-09-489-039A-10294  
 Sequence 10294, Application US/09489039A  
 Patent No. 6610836  
 GENERAL INFORMATION:  
 APPLICANT: Gary Bretton et. al  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 2709.2004001  
 CURRENT APPLICATION NUMBER: US/09/489, 039A  
 CURRENT FILING DATE: 2000-01-27  
 PRIOR APPLICATION NUMBER: US 60/117,747  
 PRIOR FILING DATE: 1999-01-29  
 NUMBER OF SEQ ID NOS: 14342  
 SEQ ID NO 10294  
 LENGTH: 348  
 TYPE: PRT

ORGANISM: *Klebsiella pneumoniae*  
US-09-489-039A-10294

Query Match	6.9%	Score 97	DB 2	Length 348
Best Local Similarity	22.7%	Pred. No. 0.067		
Matches	70	Conservative	42	Mismatches 109
			Indels	88
			Gaps	17
QY	14	IDGSKPAVQALMAVDEASRDIPRLRLVLAIEDDPGY-----AAHG---AAARKL	61	
Db	72	LDGGQVAA-----AAAGCDV---IVHANVP--PGYRHWROQLPWLHNTLLDAARQR	118	
QY	62	AA-----AENAVRYAFTVAEADRP--KVEVEITQERVYTSILRASAAAALYCV	109	
Db	119	ALVVLPGTVYNYGPN--FLIEEAQOPTRGALRVAMELEIDYVGQGRLLIVRA	176	
QY	110	GAIGVHHFRPERVEST--AAALALSAQCPVAVRPHRPIGRDAAMIVEADGSSDIGVL	167	
Db	177	G-----DFGGRAGANNMFQGLVPRGQLPRILISPTIGVGHQVAML-----PDVAAT	224	
QY	168	LGAVMABEARLDSFVRVVTCKROSGVGD--TGDDVVASLDRLMARQPR-----YDPRVQY	221	
Db	225	IALLLARHLEPRARF--HMQGHMDPDGSEMSQALQRYVAVRFGRAVVKSPFWMLVKL	281	
QY	222	AAV-----HGELDYLAGLGRSVAMVVLASDQGHVEQLVQAP	259	
Db	282	AAFPNATLRKVENHMYLMRLPVRLRNKLVDF--LGAEPHTPLDSDAVYQY--LOGLGCLP	337	
QY	260	GNAYLQEAG	268	
Db	338	AGAINQDEAG	346	

Search completed: March 23, 2006, 05:30:09  
Job time : 20.1287 secs

**THIS PAGE BLANK (USPTO)**

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 06:14:33 ; Search time 58.6719 Seconds  
(without alignments)  
1986.885 Million cell updates/sec

Title: US-10-617-038-21

Perfect score: 1110  
Sequence: 1 MNSHKRPSIVVGIDSKPA.....GNVLAQAGCTLLVVGQYTL 279

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_Main:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1410	100.0	279	4	US-10-617-038-21
2	431.5	30.6	268	4	US-10-617-038-42
3	416	29.5	272	4	US-10-617-038-27
4	214	15.2	297	4	US-10-138-473-34
5	214	15.2	297	5	US-10-450-726-4
6	208	14.8	295	4	US-10-156-761-8839
7	207	14.7	301	3	US-09-738-626-6745
8	207	14.7	301	3	US-09-746-660A-6
9	207	14.7	301	3	US-09-746-660A-20
10	207	14.7	301	4	US-10-450-055-8
11	207	14.7	301	6	US-11-006-098-114
12	206.5	14.6	290	4	US-10-156-761-8814
13	180	12.8	300	4	US-10-282-122A-53861
14	131	9.3	181	4	US-10-437-963-12675
15	127.5	9.0	3227	5	US-10-760-493-33
16	118.5	8.4	2539	4	US-10-282-122A-61655
17	113	8.0	7068	4	US-10-203-295-30
18	113	8.0	9477	4	US-10-203-295-37
19	112	7.9	11088	4	US-10-203-295-7
20	112	7.9	11096	5	US-10-732-923-20557
21	109.5	7.8	10917	5	US-10-732-923-20606
22	109	7.7	2365	4	US-10-156-761-14890
23	108	7.7	1050	4	US-10-156-761-8816
24	107.5	7.6	801	4	US-10-365-483-8838
25	106.5	7.6	860	5	US-10-925-357-22
26	106	7.5	1641	3	US-09-980-217-30
27	106	7.5	3753	3	US-09-980-217-29

28	105.5	7.5	152	4	US-10-282-122A-45977	Sequence 45977, A
29	105.5	7.5	471	4	US-10-437-963-156094	Sequence 156094, A
30	105.5	7.5	1976	5	US-10-937-730A-7	Sequence 7, Appl1
31	105.5	7.5	6291	4	US-10-329-079-41	Sequence 41, Appl1
32	103.5	7.3	533	4	US-10-156-761-13934	Sequence 13934, A
33	103	7.3	181	4	US-10-425-115-230179	Sequence 230179, A
34	103	7.3	409	4	US-10-156-761-13624	Sequence 13624, A
35	103	7.3	3025	3	US-09-980-217-19	Sequence 19, Appl1
36	103	7.3	5435	4	US-10-203-295-38	Sequence 38, Appl1
37	103	7.3	5435	5	US-10-732-923-20558	Sequence 20558, A
38	103	7.3	7429	5	US-10-937-730A-5	Sequence 5, Appl1
39	102.5	7.3	1657	5	US-10-760-493-27	Sequence 27, Appl1
40	102.5	7.3	10625	5	US-10-819-386A-3	Sequence 3, Appl1
41	102	7.2	890	5	US-10-732-923-8067	Sequence 8067, Ap
42	101.5	7.2	391	4	US-10-344-738-31	Sequence 31, Appl1
43	101.5	7.2	402	4	US-10-369-493-22935	Sequence 22935, A
44	101.5	7.2	795	4	US-10-314-657-43	Sequence 43, Appl1
45	101.5	7.2	795	5	US-10-473-193-43	Sequence 43, Appl1

## ALIGNMENTS

RESULT 1  
US-10-617-038-21  
; Sequence 21, Application US/10617038  
; Publication No. US20040057963A1  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Peter  
; APPLICANT: Rosenkrands, Ida  
; APPLICANT: Strjhn, Anette  
; TITLE OF INVENTION: Therapeutic TB Vaccine  
; FILE REFERENCE: SSI5AUSA  
; CURRENT APPLICATION NUMBER: US/10/617, 038  
; CURRENT FILING DATE: 2003-07-11  
; PRIOR APPLICATION NUMBER: DK PA 2002 01098  
; PRIOR FILING DATE: 2002-07-13  
; PRIOR APPLICATION NUMBER: US 60/401, 725  
; PRIOR FILING DATE: 2002-08-07  
; NUMBER OF SEQ ID NOS: 187  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 21  
; LENGTH: 279  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-617-038-21

Query Match 100.0%; Score 1410; DB 4; Length 279;  
Best Local Similarity 100.0%; Pred. No. 4.8e-132;  
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MNSHKRPSIVVGIDSKPAVOALMVDRAASDPIRLLYAIEPDPCGAHGAARK	60
DB	1	MNSHKRPSIVVGIDSKPAVOALMVDRAASDPIRLLYAIEPDPCGAHGAARK	60
QY	61	LAALANVRAFAFVAEADRPVKEVEITTOERPTSLIRASAAALVCVGAIGVHFRPE	120
DB	61	LAALANVRAFAFVAEADRPVKEVEITTOERPTSLIRASAAALVCVGAIGVHFRPE	120
QY	121	RVGSTAALALSAQCFAIVRPHRPVIGRDAAMIVVEADSSDGIYLGAVMAEARLDS	180
DB	121	RVGSTAALALSAQCFAIVRPHRPVIGRDAAMIVVEADSSDGIYLGAVMAEARLDS	180
QY	181	PVRVVTROGSGVDTGDDVVASLDRWLARWQPRPDVRSAAVHGSLDYLGLGRSVH	240
DB	181	PVRVVTROGSGVDTGDDVVASLDRWLARWQPRPDVRSAAVHGSLDYLGLGRSVH	240
QY	241	MVVLASDQEHVEQLVGAPEGNAVLOEAGCTLLVVGQYTL	279
DB	241	MVVLASDQEHVEQLVGAPEGNAVLOEAGCTLLVVGQYTL	279

RESULT 2

US-10-617-038-42  
; Sequence 42, Application US/10617038  
; Publication No. US20040057963A1  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Peter  
; APPLICANT: Rosenkrands, Ida  
; APPLICANT: Stryhn, Anette  
; TITLE OF INVENTION: Therapeutic TB Vaccine  
; FILE REFERENCE: SS15AUSA  
; CURRENT APPLICATION NUMBER: US/10/617,038  
; CURRENT FILING DATE: 2003-07-11  
; PRIOR APPLICATION NUMBER: DK PA 2002 01098  
; PRIOR FILING DATE: 2002-07-13  
; PRIOR APPLICATION NUMBER: US 60/401,725  
; PRIOR FILING DATE: 2002-08-07  
; NUMBER OF SEQ ID NOS: 187  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 42  
; LENGTH: 268  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-617-038-42

Query Match 30.6%; Score 431.5; DB 4; Length 268;  
Best Local Similarity 40.4%; Pred. No. 2.2e-34;  
Matches 11; Conservative 34; Mismatches 105; Indels 25; Gaps 8;

Qy 1 MNQSHKPSIVGIDGSKPAVQALMAVDEASRDIPRLLYAIEPDDPGYAAGAAARK 60  
Db 1 MSDRPARAVVVGIDGSRRAATHALMAVDEAVNDIPRLLYVVDPSGLSAAGGCGQ-- 58  
Qy 61 LAAENAVRYAFTVAEADRPVKEVEITQERRPVTSLIRASAAALVCVGAIGHFRPE 120  
Db 59 -SARARALHDSARKVEATGCPVKIETEVLCGRPLTKMQESRSAAMLCVGSGVDHYRGR 117  
Qy 121 RVGSTAALALSAQCPVAIYRPH--RVPIGDAAIVIYVDEAGSSDGVTLGAVVAEARLR 178  
Db 118 R-GVAATTLGSAALCPVAIVHPSPAPATTSQSAVAAEVDN---GVVLAHAFEEARLR 172  
Qy 179 DSPRVVTCRQSGVGDGDDV-----RASLDRLRMQRYPDPVQSAVHGEILDY 231  
Db 173 GVPRAVAHV---AAETPDDVEGSRRLAHVLSRLAHMTLLYEAVDRALJAGSACRH 225  
Qy 232 LAGIGRSVHNVLSASDQEHVEQDVGAA--PGNAV 264  
Db 230 LAANAKPGQLFV---ADSHSAHELCGAYQPGCAVL 261

RESULT 3  
US-10-617-038-27  
; Sequence 27, Application US/10617038  
; Publication No. US20040057963A1  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Peter  
; APPLICANT: Rosenkrands, Ida  
; APPLICANT: Stryhn, Anette  
; TITLE OF INVENTION: Therapeutic TB Vaccine  
; FILE REFERENCE: SS15AUSA  
; CURRENT APPLICATION NUMBER: US/10/617,038  
; CURRENT FILING DATE: 2003-07-11  
; PRIOR APPLICATION NUMBER: DK PA 2002 01098  
; PRIOR FILING DATE: 2002-07-13  
; PRIOR APPLICATION NUMBER: US 60/401,725  
; PRIOR FILING DATE: 2002-08-07  
; NUMBER OF SEQ ID NOS: 187  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 27  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-617-038-27

Query Match 29.5%; Score 416; DB 4; Length 272;

Best Local Similarity 36.8%; Pred. No. 8e-33;  
Matches 98; Conservative 46; Mismatches 116; Indels 6; Gaps 2;  
Qy 9 SIVGIDGSKPAVQALMAVDEASRDIPRLLYAIEPDDPGYAAGAAARKLAAMENAV 68  
Db 11 TIIVGIDGSHAAITPALMGVDEAISRAVPLRLVSVIKFTHPSDDVD--RDLAHARSL 67  
Qy 69 RYAFVAEADRPVKEVEITQERRPVTSLIRASAAALVCVGAIGHFRPERVGTAAA 128  
Db 68 REAQSAAVEAAGKVLKTIETDIPRGPAFVLVEASRDAMICVGSVGIGRYASSILGSTATE 127  
Qy 129 LALSACCPVAIYRPHRPIGRDAAIVIYVDEAGSSDGVTLGAVVAEARLSDPVRVTCR 188  
Db 128 LAEKACPVAVMSKVDOPASDINWIVRMTDAPDEAVLEVAAREKRLQAPLALAGR 187  
Qy 189 QSGVDTGDDVRASLDRLRMQRYPDPVQSAVHGEILDYLAGIGRSVHNVLSASD 248  
Db 188 PEELEKRIPD---GEFERVQDMHHRHEDVAVVYPTTHGTGLARFLADDERVOLAVIOGGE 244  
Qy 249 QSHVEQLVGAPGNAVLOEAGCTLLVV 274  
Db 245 AGQLARLVGPSGHPVFRHAECVLLVV 270

RESULT 4  
US-10-138-473-34  
; Sequence 34, Application US/10138473  
; Publication No. US20030165525A1  
; GENERAL INFORMATION:  
; APPLICANT: ANDERSEN, Peter  
; APPLICANT: WELDLING, Karin  
; APPLICANT: HANSEN, Christina Veggerby  
; APPLICANT: FLORIO, Walter  
; APPLICANT: OKKELS, Li Mei Meng  
; APPLICANT: SKJORT, Rikke Louise Vinther  
; APPLICANT: ROSENKRANDS, Ida  
; APPLICANT: RASMUSSEN, Peter Birk  
; TITLE OF INVENTION: TB Diagnostic Based On Antigens From M. tuberculosis  
; FILE REFERENCE: 0459-0710P  
; CURRENT APPLICATION NUMBER: US/10/138,473  
; CURRENT FILING DATE: 2002-05-02  
; PRIOR APPLICATION NUMBER: US 10/060,428  
; PRIOR FILING DATE: 2002-01-29  
; PRIOR APPLICATION NUMBER: US 09/415,884  
; PRIOR FILING DATE: 1999-10-08  
; PRIOR APPLICATION NUMBER: US 60/116,673  
; PRIOR FILING DATE: 1999-01-21  
; PRIOR APPLICATION NUMBER: DK 1998 01281  
; PRIOR FILING DATE: 1998-10-18  
; PRIOR APPLICATION NUMBER: US 60/070,488  
; PRIOR FILING DATE: 1998-01-05  
; PRIOR APPLICATION NUMBER: DK 1997 01277  
; PRIOR FILING DATE: 1997-11-10  
; PRIOR APPLICATION NUMBER: US 60/044,624  
; PRIOR FILING DATE: 1997-04-18  
; PRIOR APPLICATION NUMBER: DK 1997 00376  
; PRIOR FILING DATE: 1997-04-02  
; NUMBER OF SEQ ID NOS: 174  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 34  
; LENGTH: 297  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-138-473-34

Query Match 15.2%; Score 214; DB 4; Length 297;  
Best Local Similarity 27.2%; Pred. No. 1.3e-12;  
Matches 84; Conservative 37; Mismatches 134; Indels 54; Gaps 10;

Qy 1 MNQSHKPSIVGIDGSKPAVQALMAVDEASRDIPRLLYAIEPDDPGY 51  
Db 1 MSSGNSGLGIVGIDGSPAAQVAVRMAPADAEIRKILPLTVHNVSPVATWLEVPDPGV 60

Qy 52 AA-----HGAARKLAAENAVRFTAVEAD---RPVKVEVEITQERPVTSILIRASAA 104  
Db 61 LRWQODHG-----RHLLIDDLKVVESQASLRAGPETHSEITVPAAVPTLVDMSDA 111  
Qy 105 ALVVCAGIAGVHFRPERVSTPAALALSAOCFPAIV-----RPHRVPIGRDAAMIV 156  
Db 112 VLMVYVCGCLGSGRMWPGRLIGSVSSGLLRHAKCPVITIHEDESVMHFPOAP-----VLV 164  
Qy 157 EADSSSDIGVLLGAVMAEARLRDSPRVVTCRQSGVGDPT-----GDDVRAS-----LDR 205  
Db 165 GVDSSSASELATATAFDEASRRN--VDLVALHMSDVDSWPGIDMPATQSMAEQVLA 222  
Qy 206 WLAWMQRRYPVVRQSAAVHGEELDYLAGLGRSVHMYVLSASDQ--EHVEQLVGAPGNAV 264  
Db 223 RLAWQMQRYPVVAITRVVVRDQPAROLVQRSEEAQLVVGSGRGVAGMLVGSVGTVA 282  
Qy 265 QEAGCTLLV 273  
Db 283 QIARTPVIV 291

## RESULT 5

US-10-450-726-4  
; Sequence 4, Application US/10450726  
; Publication No. US20040242471A1  
; GENERAL INFORMATION:  
; APPLICANT: Dick, Thomas  
; APPLICANT: Boon, Calvin Ka Khin  
; TITLE OF INVENTION: DORMANCY-INDUCED MYCOBACTERIUM PROTEINS  
; FILE REFERENCE: 50316/002001  
; CURRENT APPLICATION NUMBER: US/10/450, 726  
; CURRENT FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: PCT/EP01/14551  
; PRIOR FILING DATE: 2001-12-11  
; PRIOR APPLICATION NUMBER: 0030368.5  
; PRIOR FILING DATE: 2000-12-13  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 297  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-450-726-4

Query Match 15.2%; Score 214; DB 5; Length 297;  
Best Local Similarity 27.2%; Pred. No. 1.3e-12;  
Matches 84; Conservative 37; Mismatches 134; Indels 54; Gaps 10;

Qy 1 MNDSHKPPSIYVIGIDSKPAVOALMAVDEASFDIPRLLYAIEPD-----DPGY 51  
Db 1 MSSGNSSLGIIVIGIDSDPAQVAVRMAARDALRKPLTLVHAVSPVATWLEVPFGV 60  
Qy 52 AA-----HGAARKLAAENAVRFTAVEAD---RPVKVEVEITQERPVTSILIRASAA 104  
Db 61 LRWQODHG-----RHLLIDDLKVVESQASLRAGPETHSEITVPAAVPTLVDMSKDA 111  
Qy 105 ALVVCAGIAGVHFRPERVSTPAALALSAOCFPAIV-----RPHRVPIGRDAAMIV 156  
Db 112 VLMVYVCGCLGSGRMWPGRLIGSVSSGLLRHAKCPVITIHEDESVMHFPOAP-----VLV 164  
Qy 157 EADSSSDIGVLLGAVMAEARLRDSPRVVTCRQSGVGDPT-----GDDVRAS-----LDR 205  
Db 165 GVDSSSASELATATAFDEASRRN--VDLVALHMSDVDSWPGIDMPATQSMAEQVLA 222  
Qy 206 WLAWMQRRYPVVRQSAAVHGEELDYLAGLGRSVHMYVLSASDQ--EHVEQLVGAPGNAV 264  
Db 223 RLAWQMQRYPVVAITRVVVRDQPAROLVQRSEEAQLVVGSGRGVAGMLVGSVGTVA 282  
Qy 265 QEAGCTLLV 273  
Db 283 QIARTPVIV 291

## RESULT 6

US-10-156-761-8839  
; Sequence 8839, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156, 761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 8839  
; LENGTH: 295  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-8839

Query Match 14.8%; Score 208; DB 4; Length 295;  
Best Local Similarity 28.2%; Pred. No. 5.2e-12;  
Matches 87; Conservative 39; Mismatches 117; Indels 66; Gaps 11;

Qy 9 SIIVGIDSKRPVAVOALMAVDEASRDIPLRLLYAIEP-----DDPGTAHGA 58  
Db 7 TITVIGDSRTGLDADWMAAREARRRLPLRLHAGEPVPAVPVVDVLPARTALDR 66  
Qy 59 RLKLAAMENAVRFTAVEADRPVKVEVEITQERPVTSILIRASAAALVVCAGIAGVHFR 118  
Db 67 VOLA-----YHAPALDIARR-----TETPAVALAALAAEAETLVLSRGFTGA 112  
Qy 119 PERVSTPAALALSAOCFPAIVRPHRV-----PIGRDAAM--IVENDGSSDI 164  
Db 113 GFLVGSVALAVARARARPVVLTGELAEDERPADGCTPGR--AAVLPVVLGLDLHPA 171  
Qy 165 GVLGAVMAEARLRDSPRVV-----TCRQSGVGDPTD-----DVASLDRWLAR 210  
Db 172 DRLLAVAFBAAVRASAPLVHVTWTLPTGHAAPAPLPDAAREBGARTLTTLTQPW 231  
Qy 211 QPRYPDVVRQSAAVHGEELDYLAGLGRSVHMYVLSASDQEHVEQLVGAPG-----NAV 265  
Db 232 RHKFPGTDLERVIPGHGHLRLASIRAGLIVG-----RTSAPGIGRAARSLIH 283  
Qy 266 EAGCTLLV 274  
Db 284 HAGCPVAVV 292

## RESULT 7

US-09-738-626-6745  
; Sequence 6745, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHITO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125

```

; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 6
; LENGTH: 301
; TYPE: PRF
; ORGANISM: Corynebacterium glutamicum
US-09-746-660A-6

Query Match      14.7%; Score 207, DB 3, Length 301,
Beet Local Similarity 26.1%; Pred. No. 6.7e-12;
Matches    77; Conservative   46; Mismatches 134; Indels  38; Gaps  10;

QY          10 IVVGIDGSKPAVQAALMAVDASAASDIPFLRL-----LYA---IREDDPGVAAGAA 57
           |||::|||::|||::|||::|||||::|::|::|::|::|::|::|::|::|::|::|::|
DB          7 VVAAVDGSDASKQAARWAMANTANKRGILPRLASSVTYPQFIYAEGWPPOELFDLQLAEA 66
           |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY          58 ARKLAALAENRVAFTVAEADRPVKVEITOEPRPVSLIRASAAAALVCVGAIGVHHF 117
           |:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          67 LEKINEARD-----IAIEVAPE-IKITHTIABGSPIIMLEMPDAIMIYMGGSGGLG 119
           ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY          118 RPERVESTAALLALSACFPVALIVEPHRRVPICRDAA--IVVEDAGSSDIGVLGAWMAEA 175
           :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          120 SGWMGVSVGSAVVSHAKCFVNVVRBDSA-VNEDSKYGCVVVGVDSBVQSQTATEYAFABA 178
           ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY          176 RLRF-----DSPRVYTCTQGSGVGTGDDEV-RASLDRLMLARNPP--RIPDVARVQ 220
           :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          179 EARGAELVAVHATMDMQVASLAGLAAAQQQMDEVEROOTMILERLAFLVEKYPSPTVK 238
           :||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY          221 SAHVGEILDYLTAGLGRSVMHWVLSASDOEHYE-QLVGAPGNAYLOEAGCTLTVV 274
           :||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          239 KIITRDPRVRLALEASENAQLLVGVSHGRGGFKGMLLGSTRALLOSAPCEMMVY 293
           :||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|


RESULT 9
US-09-746-660A-20
; Sequence 20, Application US/09746660A
; Publication No. US20030049804A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Krieger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zeider, Oskar
; APPLICANT: Haberhaener, Gregor
; APPLICANT: Kim, Jun-Won
; APPLICANT: Lee, Heung-Schick
; APPLICANT: Hwang, Byung-Joon
; TITLE OR INVENTION: CORNYNEBACTERIUM GLUTAMICUM GENES ENCODING
; TITLE OP INVENTION: METABOLIC PATHWAY PROTEINS
; FILE REFERENCE: BGI-121CP2
; CURRENT APPLICATION NUMBER: US/09/746, 660A
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 09/606740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 09/603124
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/142101
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: 60/148613
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/187970
; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: DE 19931420..9
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: Patentln Vers. 2.0
; SEQ ID NO 20
; LENGTH: 301
; TYPE: PRF
; ORGANISM: Corynebacterium glutamicum
US-09-746-660A-20
```





APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 8814  
LENGTH: 290  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-8814

Query Match 14.6%; Score 206.5; DB 4; Length 290;

Best Local Similarity 26.9%; Pred. No. 7.2e-12;

Matches 77; Conservative 46; Mismatches 138; Indels 25; Gaps 8;

Qy 10 IVVGIDSKPAVQALMAVDEAASRDIPRLYLAI--EPDDPGYAAG-AAARKLAAMN 66  
Db 5 LVVGVDSSESMKAVDMAADEALRGVPLRLVYLSMERYGALADGLGRPEQVTHDD 64  
Qy 67 AVRVAFTVAEADRPVKEVEITQERPVTSILIRASAAALVCVGAIGVHFRPERVSTA 126  
Db 65 IVEAAGRAHRRHADVKISTDVLPEEPAPALLREGNEAFALVLSGRSGLAELLGSVS 124  
Qy 127 AALALSAQCPRVAIYR---PHRVPIGRDAWIIVEADGSSDIGVLGAVMAARLDSYR 183  
Db 125 LAVAARAYCPVTVIRGSDNRGAGGHRRIYLAAGBEBDDPAARFAV-ABARARGCAVE 183  
Qy 184 VVTC---RSGVDPDGTGDVRS-----LDRLYARWOPRPRDVRVOSAIVHGLL 229  
Db 184 AVRWRPARRALHGLHPLSSRAARLHERATEVLEAVIS---GAPDVELRRRAVEGPAR 240  
Qy 230 DYLAGLRSVHVYVLSASDOE-HVEQLVGAEGNAVLTQAGCTLLV 274  
Db 241 KVLIDASATADLLVVGARRHGHSGQLQGRVAVHMLHNSACPRVAV 286

RESULT 13  
US-10-282-122A-53861

Sequence 53861, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangou  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haseibeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zykend, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 53861  
LENGTH: 300  
TYPE: PRT  
ORGANISM: Corynebacterium diptheriae  
US-10-282-122A-53861

Query Match 12.8%; Score 180; DB 4; Length 300;

Best Local Similarity 24.8%; Pred. No. 3.3e-09;

Matches 73; Conservative 46; Mismatches 139; Indels 36; Gaps 9;

Qy 10 IVVGIDSKPAVQALMAVDEAASRDIPRL-----LYA---IEPDDPGYAHGAA 57  
Db 7 VVAVDSSEASQNAVRAAANTANKRGVPLRLAASYTPQFLVAGVPPQELDELQSET 66  
Qy 58 ARKLAAMNVRAFTVAEADRPVKEVEITQERPVTSILIRASAAALVCVGAIGVHFR 117  
Db 67 MDHTEAR-----VVAHEVAP-DIKGYIAGSPIDMLDMSDVTMTVMGSRIGGL 119  
Qy 118 RPERVSTAALALSAQCPRVAIYR-PHRVPIGRDAWIIVEADGSSDIGVLGAVMAAR 176  
Db 120 SGWVGVSAAVYSHADCPVVVVVRSVDHVTETNKYGVVVGVGSDVSQRATEFAFEAQ 179  
Qy 177 LRDSF-VRVVTCQSGVGTGDDVRSLSRW-----LARNP---RPDVRVQS 221  
Db 180 ARGKLVAIHTWMDQVQASLAGIAAQEWETIEKEQTLLKDRQPLIERPPDVEVM 229  
Qy 222 AAHVHGLDYLAIAGRSVHVYVLSASDOE-HVEQLVGAEGNAVLTQAGCTLLV 274  
Db 240 VTRDRPVRLIEDCAHNAQLLVVSGHSGGFRGMLGSTRALLQSAFCRMVV 293

RESULT 14  
US-10-437-963-126675

Sequence 126675, Application US/10437963

Publication No. US20040123343A1

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 126675  
LENGTH: 181  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURES:

OTHER INFORMATION: Clone ID: PAT\_MRT4530\_29199C.1.pep  
US-10-437-963-126675

Query Match

Best Local Similarity

9.3%; Score 131; DB 4; Length 181;

30.4%; Pred. No. 0.00013;

Matches 51; Conservative 23; Mismatches 54; Indels 40; Gaps 7;

Job time : 60.6719 BECV

```

QY 10 IVVGIDSSKRAVQAALMAVD-----EASSRDTL-RLTYALE-----P 46
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 VLAAVDDSSKSHALSWYDLHLFPFAALATGDCGEEQVPRPAFLVYTHALEPLHYMFP 62
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 47 DDPRYAHGAA-----ARKLAAENAVRYAFTAVEADR-----PVKVEVEITQERPVTS 96
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 VGGESAVYGAASMMELVRAQAQENAMNLLVRRKLLCERRGVAAALVAVEGE-----PREA 117
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 97 LIRAS--AAALYCVGALGVHFRPRPVGSTAAALATLSNOCPVALVR 142
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 118 LCPAAEDAGKGLLVGSRGLAKRAFLGVSVDYCNHRASCPIMVKR 165

```

### RESULT 15

```

US-10-760-493-33
: Sequence 33, Application US/10760493
: Publication No. US20050187167A1
: GENERAL INFORMATION:
: APPLICANT: Ecopia Biosciences Inc
: APPLICANT: Farnet, Chris M.
: APPLICANT: McAlpine, James B.
: APPLICANT: Zazopoulos, Emmanuel
: APPLICANT: Bachmann, Brian O.
: APPLICANT: Pitraee, Mahmood
: TITLE OF INVENTION: POLYMER POLYKETIDES, PROCESSES FOR THEIR PRODUCTION AND THEIR USE
: FILE REFERENCE: 3004-9US
: CURRENT APPLICATION NUMBER: US/10/760,493
: CURRENT FILING DATE: 2004-01-21
: PRIOR APPLICATION NUMBER: USSN 60/441,123
: PRIOR FILING DATE: 2003-01-21
: PRIOR APPLICATION NUMBER: USSN 60/469,810
: PRIOR FILING DATE: 2003-05-13
: PRIOR APPLICATION NUMBER: USSN 60/491,516
: PRIOR FILING DATE: 2003-08-01
: PRIOR APPLICATION NUMBER: USSN 60/494,568
: PRIOR FILING DATE: 2003-08-13
: NUMBER OF SEQ ID NOS: 78
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 33
: LENGTH: 3227
: TYPE: PR1
: ORGANISM: Streptomyces aizunensis
US-10-760-493-33

```

Query Match	9.0%;	Score 127.5;	DB 5;	Length 3227;	:	
Best Local Similarity	27.6%;	Pred. No. 0.015;			:	
Matches	81;	Conservative	23;	Mismatches 115;	Indels 75;	Gaps 14;

QY	41	LYALIEPDDPGVAALGAARAKLAAAMENAVAFVAVEADR	-----PVAVE-----	85
Db	2592	LVVVPENGPDEMTGAVLRVLADRGAEVRYTVVPADGDADRDRLAATLAAETIDGAAPAGVL		2655
QY	86	-----VEITGERPVTSLIRASA-----	AAALVCV--GAIGVHHFRERVGSTPA	127
Db	2652	SLMLAVESAEIRHTTGLATTAALVQALGDADVAAPLWCVRGAHSV--	ARTERLQDBAQ	2709
QY	128	ALALSAOCFVAIVRPHR-----	VGIRGDAMI-----VNEADSSD-----	IGVLQAV 171
Db	2710	ALVSGFGRVAVLEYPDRNGGLVLDLEPADGRTLERLAGVLADGSGEDVALRASGLFERR		2765
QY	172	MAEARLARDPYRVVYTCROSG-----	VGDITGDDVRASLIDRMILAR-----	NOPRYPDVRV 219
Db	2770	LVMHPLADLT-AAVREMRPGFTLVYGGTG-ALGVAHVAEMLENGAEHLLTSSRRGPD---		2822
QY	220	QSAAVHGGILDYLAGLGRSVYHNVVLSADQDEHYEBDVGCA-----	PGNAVLDQAG	268
Db	2825	--ARGAAELREGLTALGAQVTTATCTMDMDRDVAVALIAAVPADQGLTVVMHTAG		2876

Search completed: March 23, 2006, 06:28:09

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 06:17:23 ; Search time 6.6953 Seconds  
(without alignments)  
1192.746 Million cell updates/sec

Title: US-10-617-038-21

Perfect score: 1410  
Sequence: 1 MNSGKPSIVIGIDSKPA.....GNAVLOEAGCTLVVGOQYL 279

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 169630 seqs, 2862289 residues

Total number of hits satisfying chosen parameters: 169630

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_New:  
1: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	209	14.8	305	US-11-087-099-3468	Sequence 3468, Ap
2	207	14.7	301	US-11-055-822-14	Sequence 14, Appl
3	136	9.6	152	US-11-087-099-6347	Sequence 6347, Ap
4	99.5	7.1	828	US-11-087-099-6963	Sequence 6963, Ap
5	98.5	7.0	171	US-11-096-568A-32572	Sequence 32572, A
6	98.5	7.0	179	US-11-096-568A-27882	Sequence 27882, A
7	97	6.9	357	US-11-096-568A-17171	Sequence 17171, A
8	97	6.9	370	US-11-096-568A-17170	Sequence 17170, A
9	95.5	6.8	316	US-11-096-568A-13340	Sequence 13340, A
10	94.5	6.7	345	US-11-096-568A-18547	Sequence 18547, A
11	93.5	6.6	161	US-11-087-099-2546	Sequence 2546, Ap
12	93.5	6.6	161	US-11-087-099-6989	Sequence 6989, Ap
13	93.5	6.6	333	US-11-096-568A-18548	Sequence 18548, A
14	91	6.5	174	US-11-096-568A-27471	Sequence 27471, A
15	90.5	6.4	174	US-10-454-437-254	Sequence 254, Ap
16	90.5	6.4	1086	US-11-052-554A-380	Sequence 380, App
17	90.5	6.4	3689	US-11-075-185-4	Sequence 4, Appl1
18	90	6.4	592	US-10-840-688-22	Sequence 22, Appl1
19	90	6.4	7968	US-11-143-980-49	Sequence 49, Appl1
20	90	6.4	8695	US-11-205-109-15	Sequence 15, Appl1
21	89	6.3	268	US-11-096-568A-18549	Sequence 18549, A
22	88.5	6.3	159	US-11-087-099-12270	Sequence 12270, A
23	88	6.2	7102	US-11-143-980-48	Sequence 48, Appl1
24	87	6.2	150	US-11-087-099-3411	Sequence 3411, Ap
25	87	6.2	162	US-11-087-099-5731	Sequence 5731, Ap

26	86.5	6.1	920	US-11-087-099-11404	Sequence 11404, A
27	86	6.1	152	US-11-096-568A-4792	Sequence 4792, Ap
28	86	6.1	163	US-11-087-099-11409	Sequence 11409, A
29	86	6.1	163	US-11-096-568A-4791	Sequence 4791, Ap
30	86	6.1	164	US-11-087-099-6603	Sequence 6603, Ap
31	86	6.1	165	US-11-096-568A-4790	Sequence 4790, Ap
32	85.5	6.1	181	US-11-087-099-10726	Sequence 10726, A
33	85.5	6.1	852	US-10-645-441-15	Sequence 15, Appl1
34	85.5	6.1	852	US-10-725-475-7	Sequence 7, Appl1
35	85.5	6.1	852	US-11-050-804-6	Sequence 6, Appl1
36	85.5	6.1	3300	US-11-052-554A-133	Sequence 133, App
37	85	6.0	316	US-11-096-568A-16455	Sequence 16455, A
38	85	6.0	329	US-11-087-099-8594	Sequence 8594, Ap
39	85	6.0	356	US-11-096-568A-16454	Sequence 16454, A
40	85	6.0	376	US-11-096-568A-16453	Sequence 16453, A
41	84.5	6.0	3655	US-11-075-185-5	Sequence 5, Appl1
42	84	6.0	954	US-11-087-099-10843	Sequence 10843, A
43	83.5	5.9	137	US-11-087-099-507	Sequence 507, App
44	83.5	5.9	137	US-11-087-099-508	Sequence 508, App
45	83.5	5.9	247	US-11-096-568A-12417	Sequence 12417, A

## ALIGNMENTS

RESULT 1  
US-11-087-099-3468  
; Sequence 3468, Application US/11087099  
; Publication No. US20060041961A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53450)B EP  
; CURRENT APPLICATION NUMBER: US/11/087,099  
; CURRENT FILING DATE: 2005-03-22  
; NUMBER OF SEQ ID NOS: 12464  
; SEQ ID NO 3468  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Thermobifida fusca  
US-11-087-099-3468

Query Match 14.8%; Score 209; DB 7; Length 305;  
Best Local Similarity 28.0%; Pred. No. 8.6e-12;  
Matches 84; Conservative 43; Mismatches 133; Indels 40; Gaps 10;

QY 7 PPSIVIGIDSKPAVQALMAVDEAASRDIPRLIYA-----IEPDDGAYAHGAAR 59  
DB 8 PGTIVVGVDSEGLHLDWAIDAAAGGATRLIVYMGDLPLVTPLGGPRTAPSEVS 67  
QY 60 KLAAG--ENAVRYAFVAVEAADRPVKVEITQERPTSLIRASAAALVCVGAIGHNF 117  
DB 68 QAAALLLEALR-----RVQDAAPSLRAVTEVSRAEHALIKSKQDELIVSGRYSGV 123  
QY 118 RPERVSTAAALALSAQCPVAIVRPHRVPIGRDAWITVEADSSDIGVLGAVMAA-R 176  
DB 124 ASLEFGVAAQVAVSHATCPVVVPTQEAARGRVAVGVGDSEHAALALRFALVEAR 183  
QY 177 LR-----DSVRYVVTGKSGVGTGDDVRSLSRMLARM--QRRYP-----DV 217  
DB 184 LRRELVAIVVMAQAPDAVDEPFTVLQADVAVDREQVARRAEWMLRTVDEARTPLTORPV 243  
QY 218 RVQSAANH--GELIDYLAGRSVHMVVLASDQ--EHVEDLVGAPGNAVLOEAGCTLVV 274  
DB 244 RVEPPEKPAALALDE---GADADLVGSRGRGFTGLILSGVQVLAHNAVVPAAVV 299

RESULT 2  
US-11-055-822-14  
; Sequence 14, Application US/11055822  
; Publication No. US20050260707A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompejus, Markus



```

; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 32572
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(171)
; OTHER INFORMATION: Ceres Seq. ID no. 13593512
US-11-096-568A-32572

```

```

Query Match          7.0%; Score 98.5; DB 7; Length 171;
Best Local Similarity 23.6%; Pred. No. 0.06;
Matches 37; Conservative 31; Mismatches 64; Indels 25; Gaps 6;

```

```

QY 10 IVGIDGSKPAVQAALMAVD---EASRDIPRLLYAIEPD-----DEGYAHG-- 55
DB 10 VVAVADDSSEVSMELRWALDNLKLSSSSDSFVYLH-VQSPSPVAAGVSPGTIPFGGPS 68
QY 56 ---AAAKLAANAAYVAFVAVE-----AADPVKVEVETQERPTSLIRA--SAAAA 105
DB 69 GLEVPARTALIEQHQKRTDTILHSAQICEKSVNKTQVIGDPRYKICEAVENLHAD 128
QY 106 LVCGAIGVHHFRPERVGTSTAALALSAQCPVAIVRP 142
DB 129 LVMGSRAYGRIRKMPFGSVSNCTNMAHCPVVIKRP 165

```

```

RESULT 6
US-11-096-568A-27882
; Sequence 27882, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 27882
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(179)
; OTHER INFORMATION: Ceres Seq. ID no. 2123864
US-11-096-568A-27882

```

```

Query Match          7.0%; Score 98.5; DB 7; Length 179;
Best Local Similarity 23.0%; Pred. No. 0.064;
Matches 38; Conservative 30; Mismatches 64; Indels 33; Gaps 6;

```

```

QY 10 IVGIDGSKPAVQAALMAVD---EASRDIPRLLYAIEPD-----DEGYAHGAA- 57
DB 10 VVAVADDSSEVSMELRWALDNLKLSSSSDSFVYLH-VQSPSPVAAGVSPGTIPFGGGL 68
QY 58 ---ARKLAANAAYVAFVAVEADR-----PVKVEVETQERPTSLIR 99
DB 69 EVRFPTAIEHQKRTDTILHSAQICEKSVNKTQVIGDPRYKICEAVENLHAD 128
QY 100 A--SAAALVCGAIGVHHFRPERVGTSTAALALSAQCPVAIVRP 142
DB 129 AVENTHADLVMGSRAYGRIRKMPFGSVSNCTNMAHCPVVIKRP 173

```

```

RESULT 7
US-11-096-568A-17171
; Sequence 17171, Application US/11096568A

```

```

; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 17171
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(357)
; OTHER INFORMATION: Ceres Seq. ID no. 12356312
US-11-096-568A-17171

```

```

Query Match          6.9%; Score 97; DB 7; Length 357;
Best Local Similarity 25.9%; Pred. No. 0.22;
Matches 67; Conservative 27; Mismatches 93; Indels 72; Gaps 14;

```

```

QY 6 KPFSIVGIDGSKPAVQAALMAVDEASRDIPRLLYAIA-----EPDDPGY-- 51
DB 20 KPPLPLSVRLQAGLTAAIDAVER--RDGTNRCLYGVIDRLLSARASPRPDASGVS 76
QY 52 -----AAAGAAARKLAANAAYVAFVAVEADRPEVKEVETQERPTSLIRASAAA 105
DB 77 YDVTMDASRGIMARVAPRPA-----AADRL-----PVVYFHGGGFA- 113
QY 106 LVCGAIGVHHFRPERVGTSTAALALSAQCPVAIVRHRPVIGRDAWIVVADGSSDIG 165
DB 114 -LSPAIQFPNGVCRRICALLGAVVSVNRYLA--PEHRPAAVYD-----DGVDALR 162
QY 166 VLLGAVMAER-----LRDS-PYRVVTCRSGVGTGDDVRAISLDRKTLARWQRYPRVY 219
DB 163 FL-----DARGGVPGIDDPVLDGTCFLAGSAGGNIVHAVNMAAAMPASARALRV 216
QY 220 QSAVHGELDYLAGRS 238
DB 217 --AGVF-PVQPIFGGVERT 232

```

```

RESULT 8
US-11-096-568A-17170
; Sequence 17170, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 17170
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(370)
; OTHER INFORMATION: Ceres Seq. ID no. 12356311
US-11-096-568A-17170

```

```

Query Match          6.9%; Score 97; DB 7; Length 370;
Best Local Similarity 25.9%; Pred. No. 0.23;
Matches 67; Conservative 27; Mismatches 93; Indels 72; Gaps 14;

```

```

QY 6 KPFSIVGIDGSKPAVQAALMAVDEASRDIPRLLYAIA-----EPDDPGY-- 51
DB 33 KPPLPLSVRLQAGLTAAIDAVER--RDGTNRCLYGVIDRLLSARASPRPDASGVS 89

```





Db 119 AMLVVGSHGGAIKRAVLGVSVDYCAHHAHCTVMIVKPKI 159

RESULT 12

US-11-087-099-6989  
 ; Sequence 6989, Application US/11087099  
 ; Publication No. US20060041961A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Abad, Mark S. et al.  
 ; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
 ; FILE REFERENCE: 38-21(53450)B EP  
 ; CURRENT APPLICATION NUMBER: US/11/087.099  
 ; CURRENT FILING DATE: 2005-03-22  
 ; NUMBER OF SEQ ID NOS: 12464  
 ; SEQ ID NO 6989  
 ; LENGTH: 161  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 US-11-087-099-6989

Query Match 6.6%; Score 93.5; DB 7; Length 161;  
 Best Local Similarity 26.7%; Pred. No. 0.16;  
 Matches 43; Conservative 25; Mismatches 70; Indels 23; Gaps 6;

Qy 2 NQSHKPSIVVIGIDSKPAVOALMAVDEAASRDIPRLTYAIEP---DDPGYAAHGAA- 57

Db 5 SSSSEKQVVLGIDDSBSTALNKLHDFPSPFVKLVLSRPTATSAVGFAGPGAAE 64

Qy 58 -----ARKLAA--AENAVRYAFTAVEAADRPVK--VEVITQERPTSLIRA--SAAA 104

Db 65 ILPIVSDRLKRIAAVLETAQQLCF-----NKSVDVTAEVDEDPNNVLCDAVDKYRA 118

Qy 105 ALVCGAIGVHHFRPERVGSTAALALSAQCPVAIVRHRV 145

Db 119 AILVVGSHGGAIKRAVLGVSVDYCAHHAHCTVMIVKPKI 159

RESULT 13

US-11-096-568A-18548  
 ; Sequence 18548, Application US/11096568A  
 ; Publication No. US20060048240A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alexandrov, Nickolai et al.  
 ; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
 ; FILE REFERENCE: 2750-1592PUS2  
 ; CURRENT APPLICATION NUMBER: US/11/096.568A  
 ; CURRENT FILING DATE: 2005-04-01  
 ; NUMBER OF SEQ ID NOS: 34471  
 ; SEQ ID NO 18548  
 ; LENGTH: 333  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays subsp. mays  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)\_(333)  
 ; OTHER INFORMATION: Cereb Seq. ID no. 12365697  
 US-11-096-568A-18548

Query Match 6.6%; Score 93.5; DB 7; Length 333;  
 Best Local Similarity 21.9%; Pred. No. 0.42;  
 Matches 73; Conservative 27; Mismatches 96; Indels 137; Gaps 15;

Qy 29 DEASRDIPRLTYAIEPDDPG--YAAHGAARKLAALMAVRYAFTAVEAADRPVKYEV 86

Db 10 DASAHGSSP--LHPHQPLSPSPSPFFSQAVA--VAAASYGANYRFAAPDEAGQLSPSQM 66

Qy 87 EITQER-----PVTSLIRASAAA-----ALVCGAIG----- 113

Db 67 QYIARLHLQRRQQTSVLGFPAQPMKASASAAAPARAPQKLYPRRAAAALGGVGGGDP 126

Qy 114 -----VHHFR----- 118

Db 127 PAQPHPALRLRRRGGGAGLRPGRPPRRRGAQPLPRQRLLPRAPRLRRQAPEPV 186

Qy 119 PE--RVGSTAAALALSAQCPVAIVRHR-----VPIGRDAAM 153

Db 187 PEHRRVQEGROEALRSRCV--VVRPHQQLLLAVRRRDLVLRVRRHRLVPPVHVAV 245

Qy 154 IVEADGSSD--IGVLGAVMAEARLDSPVYVTCRQSGVGDGDDVRASTL--DRWLAR 209

Db 246 RLARADAGDAALQGAQVGRRLRAHQVPVVRDR-----IGLPRQLIVAPCWLIG 299

Qy 210 WQPRYPDVVRQSA-----VHGE-----LLDYIAG 234

Db 300 WLPSPYHCRRAAGGGLKKNVGRRADGILDILIG 332

RESULT 14

US-11-096-568A-27471  
 ; Sequence 27471, Application US/11096568A  
 ; Publication No. US20060048240A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alexandrov, Nickolai et al.  
 ; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
 ; FILE REFERENCE: 2750-1592PUS2  
 ; CURRENT APPLICATION NUMBER: US/11/096.568A  
 ; CURRENT FILING DATE: 2005-04-01  
 ; NUMBER OF SEQ ID NOS: 34471  
 ; SEQ ID NO 27471  
 ; LENGTH: 174  
 ; TYPE: PRT  
 ; ORGANISM: Arabidopsis thaliana  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)\_(174)  
 ; OTHER INFORMATION: Cereb Seq. ID no. 1817050  
 US-11-096-568A-27471

Query Match 6.5%; Score 91; DB 7; Length 174;  
 Best Local Similarity 23.1%; Pred. No. 0.3;  
 Matches 37; Conservative 30; Mismatches 65; Indels 28; Gaps 6;

Qy 10 IVVGIDSKPAVOALMAVD-----EASRDIPRLTYAIEPD-----DPGYAAG--- 55

Db 10 VVAVDSDSEVSMELRALNKLKSSSSDSFVLLH--VQPSPEVAGVSGTIPFGSPS 68

Qy 56 ---AAARKLAABENAVRYAFTAVEADR-----PYKVEVITQERPTSLIRA--SA 102

Db 69 GLEVPFAFTAAIEQHKRITDTILIEHASQICAEKSVSRVNVKTQVIGDPKXKICEAVENL 128

Qy 103 AALVCGAIGVHHFRPERVGSTAALALSAQCPVAIVR 142

Db 129 HADLLVVGSRAYGRKIMPLGVSVNCTNNHACHCVVIK 168

RESULT 15

US-10-454-437-254  
 ; Sequence 254, Application US/10454437  
 ; Publication No. US20050277115A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Pompeius, Markus  
 ; APPLICANT: Krieger, Burkhard  
 ; APPLICANT: Schroder, Hartwig  
 ; APPLICANT: Zelder, Oskar  
 ; APPLICANT: Haberhauser, Gregor  
 ; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS  
 ; FILE REFERENCE: BGI-128CPCN  
 ; INVOLVED IN HOMEOSTASIS AND ADAPTATION  
 ; CURRENT APPLICATION NUMBER: US/10/454.437  
 ; CURRENT FILING DATE: 2003-06-13  
 ; PRIOR APPLICATION NUMBER: US 60/141031  
 ; PRIOR FILING DATE: 1999-06-25  
 ; PRIOR APPLICATION NUMBER: DE 19931636.8



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 04:56:29 ; Search time 90.1288 Seconds  
(without alignments)  
1652.628 Million cell updates/sec

Title: US-10-617-038-22

Perfect score: 1720  
Sequence: 1 MTEPAAWDEGKPRITLTLMN.....TEVGQDQYVHPIVPEASP 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_21: \*  
1: geneseqp1980s: \*  
2: geneseqp1990s: \*  
3: geneseqp2000s: \*  
4: geneseqp2001s: \*  
5: geneseqp2002s: \*  
6: geneseqp2003s: \*  
7: geneseqp2003bs: \*  
8: geneseqp2004s: \*  
9: geneseqp2005s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1720	100.0	339	6 ABU34644	ABU34644 Protein e
2	1720	100.0	339	6 ABU36709	ABU36709 Protein e
3	1720	100.0	339	8 ADI37301	ADI37301 M. tuberc
4	726	42.2	310	8 ADS30122	ADS30122 Bacterial
5	726	42.2	315	9 ABM95087	ABM95087 M. xanthu
6	617.5	35.9	308	8 ADS21176	ADS21176 Bacterial
7	597.5	34.7	505	6 ABU22654	ABU22654 Protein e
8	587.5	34.2	338	6 ABU20143	ABU20143 Protein e
9	534.5	31.1	310	4 AAU38983	AAU38983 Salmonell
10	531.5	30.9	309	4 AAU34587	AAU34587 E. coli c
11	531.5	30.9	309	8 ABU28646	ABU28646 Protein e
12	531.5	30.9	309	8 ADN18198	ADN18198 Bacterial
13	531.5	30.9	311	6 ABU28039	ABU28039 Protein e
14	530.5	30.3	314	7 AB063860	AB063860 Klebsiell
15	322	18.7	318	7 ADH88496	ADH88496 Enterococ
16	311	18.1	303	6 ABU17706	ABU17706 Protein e
17	304	17.7	309	6 ABU24165	ABU24165 Protein e
18	302	17.6	1249	4 ABG21666	ABG21666 Novel hum
19	299.5	17.4	340	4 AAU49793	AAU49793 Propionib
20	299.5	17.4	340	6 ABM46312	ABM46312 Propionib
21	298	17.3	321	7 ABH85922	ABH85922 Enterococ
22	294	17.1	304	6 ADB06322	ADB06322 Alloiococ
23	290	16.9	305	4 AAU35194	AAU35194 Enterococ
24	289	16.8	305	6 ABU29145	ABU29145 Protein e

25	284	16.5	306	6 ABU42743	ABU42743 Protein e
26	284	16.5	312	5 ABP40738	ABP40738 Staphyloc
27	284	16.5	312	8 ADS08044	ADS08044 Staphyloc
28	283	16.5	306	6 ADA89524	ADA89524 Staphyloc
29	283	16.5	306	6 ABU16314	ABU16314 Protein e
30	282	16.5	306	6 ABM72329	ABM72329 Staphyloc
31	282	16.4	303	4 AAU33897	AAU33897 Staphyloc
32	282	16.4	306	4 AAU36707	AAU36707 Staphyloc
33	279.5	16.2	303	9 ADB07554	ADB07554 Fructo-ol
34	279.5	16.2	312	6 ADB07554	ADB07554 Alloiococ
35	279	16.2	330	4 AAB86203	AAB86203 C. glutam
36	279	16.2	330	4 AAB79271	AAB79271 Corynebac
37	279	16.2	330	4 AAG91870	AAG91870 C. glutami
38	279	16.2	330	6 ABP60402	ABP60402 Corynebac
39	279	16.2	330	7 ADD13373	ADD13373 C. glutam
40	278.5	16.2	310	5 ABP30506	ABP30506 Streptococ
41	278.5	16.2	310	8 ADV88590	ADV88590 Streptococ
42	278.5	16.2	310	8 ADV79843	ADV79843 Streptococ
43	278.5	16.2	310	8 ADV81996	ADV81996 Streptococ
44	278.5	16.2	320	5 ABP28871	ABP28871 Streptococ
45	276.5	16.1	302	4 AAU33459	AAU33459 Enterococ

#### ALIGNMENTS

RESULT 1	
ABU34644	ABU34644 standard; protein; 339 AA.
ID	ABU34644;
AC	ABU34644;
DT	19-JUN-2003 (first entry)
XX	
XX	Protein encoded by Prokaryotic essential gene #20171.
DB	
XX	Antisense; prokaryotic essential gene; cell proliferation; drug design.
KW	Mycobacterium bovis.
OS	
XX	WO20027183-A2.
PN	
XX	03-OCT-2002.
PD	
XX	
PF	21-MAR-2002; 2002WO-US009107.
XX	
PR	21-MAR-2001; 2001US-00815242.
PR	06-SEP-2001; 2001US-00948993.
PR	25-OCT-2001; 2001US-0342923P.
PR	08-FEB-2002; 2002US-00072851.
PR	06-MAR-2002; 2002US-0362699P.
XX	
PA	(EUIT-) ELITRA PHARM INC.
XX	
PI	Wang J, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,
PT	Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;
XX	WPI; 2003-028926/02.
DR	N-PSDB; ACA38514.
XX	
PT	New antisense nucleic acids, useful for identifying proteins or screening
PT	for homologous nucleic acids required for cellular proliferation to
PT	isolate candidate molecules for rational drug discovery programs.
XX	
PS	Claim 25; SEQ ID NO 62568; 1766pp; English.
XX	
CC	The invention relates to an isolated nucleic acid comprising any one of
CC	the 6213 antisense sequences given in the specification where expression
CC	of the nucleic acid inhibits proliferation of a cell. Also included are:
CC	(1) a vector comprising a promoter operably linked to the nucleic acid
CC	encoding a polypeptide whose expression is inhibited by the antisense
CC	nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC	polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway;  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acid are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 339 AA;

Query Match 100.0%; Score 1720; DB 6; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 3,1e-167; Mismatches 0; Gaps 0;  
 Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEPAMDEGKPRITITLTMNPAIDITTSVDVVRPTEKMRGAPRYDPGGGGINVARIHV 60  
 Db 1 MTEPAMDEGKPRITITLTMNPAIDITTSVDVVRPTEKMRGAPRYDPGGGGINVARIHV 60  
 QY 61 LGGCSTALFPAGGSGTSLMALLDGAGVPPRVIPIAASTRESTVNESRTAKOYRFLPG 120  
 Db 61 LGGCSTALFPAGGSGTSLMALLDGAGVPPRVIPIAASTRESTVNESRTAKOYRFLPG 120  
 QY 121 PSLTVAEOECLELDELGAASAFAFVVASGSLPFGVADYORVADICRBSSTPLILDTSG 180  
 Db 121 PSLTVAEOECLELDELGAASAFAFVVASGSLPFGVADYORVADICRBSSTPLILDTSG 180  
 QY 181 GGLDHISGVFLKASVRELRECVGSELTTEPEOLAAAHLEIDRGRAEVVVVSLGSGCAL 240  
 Db 181 GGLDHISGVFLKASVRELRECVGSELTTEPEOLAAAHLEIDRGRAEVVVVSLGSGCAL 240  
 QY 241 LATRHASHRPSISIMPTAVSGVADAMVAATITVGLSRGWSLTKSVRLGNNAAGAMLTPG 300  
 Db 241 LATRHASHRPSISIMPTAVSGVADAMVAATITVGLSRGWSLTKSVRLGNNAAGAMLTPG 300  
 QY 301 TAAACNRDVERFELAAEPTEVGODQYVWHPITVNPASP 339  
 Db 301 TAAACNRDVERFELAAEPTEVGODQYVWHPITVNPASP 339

RESULT 2

ABU36709 standard; protein: 339 AA.

XX ABU36709;

DT 19-JUN-2003 (first entry)

DE Protein encoded by prokaryotic essential gene #22236.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Mycobacterium tuberculosis.

XX W0200277183-A2.

PD 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.  
 PF 21-MAR-2001; 2001US-00815242.  
 XX 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342973P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haeelbeck R, Ohlsen KL, Zykkind JW;  
 PI Wall D, Trivick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR N-PSDB; ACA40579.

PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 64633; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 339 AA;

Query Match 100.0%; Score 1720; DB 6; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 3,1e-167; Mismatches 0; Gaps 0;  
 Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEPAMDEGKPRITITLTMNPAIDITTSVDVVRPTEKMRGAPRYDPGGGGINVARIHV 60  
 Db 1 MTEPAMDEGKPRITITLTMNPAIDITTSVDVVRPTEKMRGAPRYDPGGGGINVARIHV 60  
 QY 61 LGGCSTALFPAGGSGTSLMALLDGAGVPPRVIPIAASTRESTVNESRTAKOYRFLPG 120  
 Db 61 LGGCSTALFPAGGSGTSLMALLDGAGVPPRVIPIAASTRESTVNESRTAKOYRFLPG 120  
 QY 121 PSLTVAEOECLELDELGAASAFAFVVASGSLPFGVADYORVADICRBSSTPLILDTSG 180  
 Db 121 PSLTVAEOECLELDELGAASAFAFVVASGSLPFGVADYORVADICRBSSTPLILDTSG 180

QY 181 GGLQHISSGVFLKASVRELRCEVSELTTEPEQLAAAHLEIDRGRAEVVVVSLGSGCAL 240  
DB 181 GGLQHISSGVFLKASVRELRCEVSELTTEPEQLAAAHLEIDRGRAEVVVVSLGSGCAL 240  
QY 241 LATTHASHRFSIIPMTAVSGVAGDAMVAATVGLSRGWSLTKSVRLGNAGAAMLLTPG 300  
DB 241 LATTHASHRFSIIPMTAVSGVAGDAMVAATVGLSRGWSLTKSVRLGNAGAAMLLTPG 300  
QY 301 TAACNRDDVERFFELAAEPTVEGDDQYVMHPIVNPPEASP 339  
DB 301 TAACNRDDVERFFELAAEPTVEGDDQYVMHPIVNPPEASP 339

RESULT 3  
AD137301  
ID AD137301 standard; protein; 339 AA.

AC AD137301;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE M. tuberculosis low oxygen induced antigen Rv2029c SEQ ID NO:22.  
XX  
KM mycobacterial infection; vaccine; tuberculosis;  
KM mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;  
KM low oxygen induced antigen.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN W0204006952-A2.  
XX  
PD 22-JAN-2004.  
XX  
PF 08-JUL-2003; 2003WO-DK000477.  
XX  
PR 13-JUL-2002; 2002DK-00001098.  
XX  
PA (STAT-) STATENS SERUM INST.  
XX  
PI Andersen P, Rosenkrande I, Stryhn A;  
XX  
DR WPI; 2004-122778/12.  
XX  
DR N-PSDB; AD137346.  
XX  
XX  
PT Use of one or more polypeptides or their fragments, which are expressed  
PT during the latent stage of the mycobacterial infection, and/or nucleic  
PT acids encoding the polypeptides, for a therapeutic vaccine against  
PT tuberculosis.  
XX  
PS Claim 3; SEQ ID NO 22; 76pp; English.  
XX  
CC The present invention describes polypeptides or their fragments, which  
CC are expressed during the latent stage of a mycobacterial infection,  
CC and/or nucleic acids encoding the polypeptides, which are useful for  
CC creating a therapeutic vaccine against tuberculosis. Also described: (1)  
CC a therapeutic vaccine against tuberculosis comprising one or more  
CC polypeptides; (2) a method for treating an animal, including a human  
CC being, with tuberculosis caused by virulent mycobacteria, e.g. by  
CC Mycobacterium tuberculosis, M. africanum or M. bovis; (3) a method for  
CC immunising an animal, including a human being, against tuberculosis  
CC caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis  
CC caused by virulent mycobacteria in an animal, including a human being;  
CC (5) a method for diagnosing previous or ongoing infection with a virulent  
CC mycobacterium; and (6) a method of diagnosing Mycobacterium tuberculosis  
CC infection in a subject. The polypeptides have antibacterial activities,  
CC and can be used in vaccines and in gene therapy. The polypeptides are  
CC useful for the manufacture of a therapeutic vaccine for treating an  
CC individual who is infected by a virulent mycobacterium, e.g. M.  
CC tuberculosis, and who is not vaccinated with BCG against tuberculosis.  
CC The present sequence represents a low oxygen induced antigen, which is  
CC used in the exemplification of the present invention.  
XX  
SQ Sequence 339 AA;

Query Match 100.0%; Score 1720; DB 8; Length 339;  
Best Local Similarity 100.0%; Pred. No. 3.1e-167;  
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEPAAWDEKPRITLTNNPALDITTSVDVVRPTEKMRGAPRYDGGGINVARIYHV 60  
DB 1 MTEPAAWDEKPRITLTNNPALDITTSVDVVRPTEKMRGAPRYDGGGINVARIYHV 60  
QY 61 LGGCTSLFPAGSGTSLMALIGDAGVPPRVPIAASTRESFTVNESRTAKOYRFVLP 120  
DB 61 LGGCTSLFPAGSGTSLMALIGDAGVPPRVPIAASTRESFTVNESRTAKOYRFVLP 120  
QY 121 PSLTVAEQECCLDELRCGAASAAFFVVASGSLPGVAADYQRYADICRSSTPLILDTSG 180  
DB 121 PSLTVAEQECCLDELRCGAASAAFFVVASGSLPGVAADYQRYADICRSSTPLILDTSG 180  
QY 181 GGLQHISSGVFLKASVRELRCEVSELTTEPEQLAAAHLEIDRGRAEVVVVSLGSGCAL 240  
DB 181 GGLQHISSGVFLKASVRELRCEVSELTTEPEQLAAAHLEIDRGRAEVVVVSLGSGCAL 240  
QY 241 LATTHASHRFSIIPMTAVSGVAGDAMVAATVGLSRGWSLTKSVRLGNAGAAMLLTPG 300  
DB 241 LATTHASHRFSIIPMTAVSGVAGDAMVAATVGLSRGWSLTKSVRLGNAGAAMLLTPG 300  
QY 301 TAACNRDDVERFFELAAEPTVEGDDQYVMHPIVNPPEASP 339  
DB 301 TAACNRDDVERFFELAAEPTVEGDDQYVMHPIVNPPEASP 339

RESULT 4  
ADS30122  
ID ADS30122 standard; protein; 310 AA.

AC ADS30122;  
XX  
DT 02-DEC-2004 (first entry)  
XX  
DE Bacterial polypeptide #19155.  
XX  
KM Recombinant DNA construct; transformed plant; improved plant property;  
KM cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
KM pathogen tolerance; pest tolerance; plant disease resistance;  
KM cell cycle pathway modification; plant growth regulator;  
KM homologous recombination; seed oil yield; protein yield; carbohydrate;  
KM nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
KM bacterial polypeptide.  
XX  
OS Bacteria.  
XX  
PN US2003233675-A1.  
XX  
PD 18-DEC-2003.  
XX  
PF 20-FEB-2003; 2003US-00369493.  
XX  
PR 21-FEB-2002; 2002US-0360039P.  
XX  
PA (CAOY/) CAO Y.  
PA (HINK/) HINKLE G J.  
PA (SLAT/) SLATER S C.  
PA (CHEN/) CHEN X.  
PA (GOLD/) GOLDMAN B S.  
XX  
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
XX  
DR WPI; 2004-061375/06.  
XX  
PT New recombinant DNA construct comprising a promoter positioned to provide  
PT for expression of a polynucleotide encoding a polypeptide from a  
PT microbial source, useful for producing plants with improved properties.  
XX  
PS Claim 1; SEQ ID NO 19155; 122pp; English.



KW pathogen tolerance; pest tolerance; plant disease resistance;  
 KW cell cycle pathway modification; plant growth regulator;  
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
 KW bacterial polypeptide.  
 XX Bacteria.  
 OS  
 PN US2003233675-A1.  
 XX  
 PD 18-DEC-2003.  
 XX  
 PF 20-FEB-2003; 2003US-00369493.  
 XX  
 PR 21-FEB-2002; 2002US-0360039P.  
 XX  
 PA (CAOY/) CAO Y.  
 PA (HINK/) HINKLE G J.  
 PA (SLAT/) SLATER S C.  
 PA (CHEN/) CHEN X.  
 PA (GOLD/) GOLDMAN B S.  
 XX  
 PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
 XX  
 DR WPI; 2004-061375/06.  
 XX  
 PT New recombinant DNA construct comprising a promoter positioned to provide  
 PT for expression of a polynucleotide encoding a polypeptide from a  
 PT microbial source, useful for producing plants with improved properties.  
 XX  
 PS Claim 1; SEQ ID NO 10209; 122pp; English.  
 XX  
 CC The invention relates to a recombinant DNA construct comprising a  
 CC promoter functional in a plant cell, where the promoter is positioned to  
 CC provide for expression of a polynucleotide encoding a polypeptide from a  
 CC microbial source. The invention also relates to a transformed plant  
 CC comprising the recombinant DNA construct and a method of producing a  
 CC transformed plant having an improved property. The plant is a crop plant  
 CC such as maize or soybean. The method of producing a transformed plant  
 CC having an improved property comprises transforming a plant with the  
 CC recombinant DNA construct and growing the transformed plant, where the  
 CC polynucleotide or polypeptide is useful for improving plant properties.  
 CC The recombinant DNA construct is useful for producing plants with  
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
 CC increased resistance to plant disease, better growth rate by modification  
 CC of the cell cycle pathway with plant growth regulators, increased rate of  
 CC homologous recombination, modified seed oil or protein yield and/or  
 CC content, improved yield by modification of carbohydrate, nitrogen or  
 CC phosphorus use and/or uptake, by modification of photosynthesis or by  
 CC providing improved plant growth and development under at least one stress  
 CC condition, improved lignin production or improved galactomannan  
 CC production. This sequence represents a bacterial polypeptide used in the  
 CC scope of the invention. Note: The sequence data for this patent did not  
 CC form part of the printed specification but was obtained in electronic  
 CC format from USPTO at seqdata.uspto.gov/sequence.html.  
 XX  
 SQ Sequence 308 AA:  
 Query Match 35.9%; Score 617.5; DB 8; Length 308;  
 Best Local Similarity 43.0%; Pred. No. 3.3e-54;  
 Matches 131; Conservative 55; Mismatches 118; Indels 1; Gaps 1;  
 QY 11 KPRITITLMNPAIDTTSVDVVRPREKRCAPYDPCGGGINVARIYHVCGGCTALFP 70  
 DB 2 KSSIFITLLTSLPAPFKSIVSVESVQPEHKLRCSPYQPGGGINVRAIKKLNGSSVALYP 61  
 QY 71 AGSGTSLMLALGDAVPRFVPIASTRESFTVNESRTAKOYFVLPGSLTVABEQ 130  
 DB 62 ICGTGNHFKALLVBEGIEQITTFQIKMTRENPFIVERSTNNQRFGRGSEIYKEBEQ 121  
 QY 131 CLDELKGAASAAAFVVASGLPPGVAAADYQYRVADICRSSTPIILDTSGGGLQH-1SSG 189

DB 122 MIAVIGDPSNDFEYFVTVSSGTPKGVDPDEFYTRLISQIILQKARLIDTSGAALKHCKEG 181  
 QY 190 VFLLKASVRELRBCVGSSELTTEPEQLAAHELIIDRGRAEVVVSLSQGLATRAHSHR 249  
 DB 182 IFLLCKPAINLSSELVIGELITTKGQEEAAMEIINSKVEILVYLSGGAGAFLSKQGIYH 241  
 QY 250 FSSITPMFVAVSGVAGDAMVAALITVGLSRGWSLKSVRFGNAGAAAMLTGTACRNDV 309  
 DB 242 VTAPPVKRSSTVAGDSMVAGMYLSRGNHYEDVYIKFGVACGTAAATMNFTELCXEDV 301  
 QY 310 ERFFE 314  
 DB 302 ERLYK 306  
 RESULT 7  
 ABU22654  
 ID ABU22654 standard; protein; 505 AA.  
 XX  
 AC ABU22654;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #8181.  
 XX  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS Burkholderia mallei.  
 XX  
 FN W0200277183-A2.  
 PD  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang J, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;  
 XX  
 DR WPI; 2003-022926/02.  
 DR N-PSDB; ACA26524.  
 XX  
 PT New antisense nucleic acids, useful for identifying proliferating or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 50578; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-regulated gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene



CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 505 AA;

Query Match 34.7%; Score 597.5; DB 6; Length 505;  
Best Local Similarity 41.9%; Pred. No. 7,66-52;  
Matches 127; Conservative 59; Mismatches 116; Indels 1; Gaps 1;

QY 12 PRITLTNNPALDITTSVDVVRPTEKMRGAPRYDPGGGGINVARIYHVLGGCSTALFPA 71  
DB 194 PEIVTLTPNPAIDVATCVERVTDTKLCRCGPARDPGGGGINVARIYHVLGGCSTALFPA 253  
QY 72 GGSGSLMALGAGPFRVYIPPIASTRESFTNVESTAKQYRFLPGSLTYAEQEOC 131  
DB 254 GGGGGLRLGLLDEAGVRAHGDIDAGETRENFSLSTSTGREFVFLGPPALAHMPRC 313  
QY 132 LDELGAASAFAFVVASGSLPPGVAADYQYRVADICRRSSTPLIDTSGGL-0HISGGV 190  
DB 314 VEALGRADASRYVWGSGLPPGMPDDCYALARASRGRTVDTSGPALAALDAGV 373  
QY 191 FLKASVRELRCEVGSSELTTEPEQLAAHEDIRGAEVVVSLGSGCALIATRHASHRF 250  
DB 374 YLVPSIGELRALTLGLPLEDDGARLAAARAIVAGGSAQIVALTIGDAGALVSHDDAVRL 433  
QY 251 SSIMTAVSGVAGADANVAATVGLSRGWSIKSVRLGNAGAMLTLPGRACNRDVE 310  
DB 434 PGVAVARSATIGADDSFVAGLVAAALNRGANVADARRHALAASASLSTGTGALTKEDIA 493  
QY 311 RFF 313  
DB 494 RIV 496

RESULT 8

ABU20143 ID ABU20143 standard; protein, 338 AA.

AC ABU20143;

DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #5670.

KM Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX *Borrelia cepacia*.

PN WO200277183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 25-OCT-2001; 2001US-00948893.

PR 08-FEB-2002; 2002US-0042923P.

PR 06-MAR-2002; 2002US-0362689P.

PA (ELIT-) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Haebeck R, Ohlsen KU, Zyskind JW;  
PI Wall D, Trawick UD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR N-PSDB; ACA24013.

XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 48067; 1766bp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 338 AA;

Query Match 34.2%; Score 587.5; DB 6; Length 338;  
Best Local Similarity 43.0%; Pred. No. 4,5e-51;  
Matches 132; Conservative 48; Mismatches 124; Indels 3; Gaps 3;

QY 14 IITLTNNPALDITTSVDVVRPTEKMRGAPRYDPGGGGINVARIYHVLGGCSTALFPAAG 73  
DB 31 IYVTLNPAVDVATVSEHIVDTHTKLCARPRRDPGGGGINVARTVHRLGGDVALYLAGG 90  
QY 74 STGSLMALGAGPFRVYIPPIASTRESFTNVESTAKQYRFLPGSLTYAEQEOCLD 133  
DB 91 PTGDVLTALAEHLBSERIRIAGETRENVCTEYTGARYRLMPGLVGTGEMRACAA 150  
QY 134 ELRGAASAFAFVVASGSLPPGVAADYQYRVADICRRSSTPLIDTSGGGLQ-HISGGVFL 192  
DB 151 RTDAMPSPRYVIVLSSLPARGADDLYATLARAKARGSVVVDAGRALQALKKGVHL 210  
QY 193 LKASVRELRCEVGSSELTTEPEQLAAHEDIRGAEVVVSLGSGCALIATRHASHRFSS 252  
DB 211 VKPSIGELSLAG-BPLDTSACRKASSELVARGQADIVALTIGARGAMVVTTRDRTLPG 269  
QY 253 IYPTAVSGVAGANVAATVGLSRGWSIKSVRLGNAGAMLTLPGRACNRDVERP 312  
DB 270 RPAAVSGTGAAGDSFVAGVWVALACGVPDDACRYALAAASAVERPITALCTRDVERI 329  
QY 313 F-ELAAE 318  
DB 330 HAEIVAG 336

```
RESULT 9
AAU38383
ID AUJ38383 standard; protein; 310 AA.
XX
AC AUJ38383;
XX
DT 14-FEB-2002 (first entry)
XX
DE Salmomella typhi cellular proliferation protein #274.
XX
KM Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.
XX
OS Salmomella typhi.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KU, Zykind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS56242.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
XX
XX Example 3; SEQ ID NO 13976; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the genes,
XX their use in the discovery of novel antibiotics, the essential genes
XX themselves and the encoded proteins. The prokaryotes used are Escherichia
XX coli, Staphylococcus aureus, Salmomella typhi, Klebsiella pneumoniae,
XX Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
XX useful for the identification of potential new targets for antibiotic
XX development. The antisense nucleic acids can also be used to identify
XX proteins used in proliferation, to express these proteins, and to obtain
XX antibodies capable of binding to the expressed proteins. The proteins can
XX be used to screen compounds in rational drug discovery programmes. The
XX antisense nucleic acid sequence is also useful to screen for homologous
XX nucleic acids which are required for cell proliferation in a wide variety
XX of organisms. The present sequence represents an essential prokaryotic
XX cellular proliferation protein. Note: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 310 AA;
XX
XX Query Match 31.1%; Score 534.5; DB 4; Length 310;
XX Best Local Similarity 41.4%; Pred. No. 1.1e-45;
XX Matches 126; Conservative 54; Mismatches 119; Indels 5; Gaps 4;
```

```
OY 13 RIITITMPADITTSUVVPTETMRCGAPRYDGGGGINVARVHVLGGSTLFPAG 72
DB 3 RIYTLTAPSDSATITPPIYEGKLRCSAPVFEPGGGGINVARIAIHGTAIFPAG 62
```

```
OY 73 GSTGSLMALIGDAGVFFRVIPIAASTRESFTVNESRTAQYRRVLPGPSLTVAEQEQL 132
DB 63 GATGEHLVALLADENVEVSTVDANKMTRONLHVHVESSGQYRFVWGATLDDDEFQ-L 121
OY 133 DELGGAASAFAVVAASGLSPGVADYYGRVADICRBSSTPLIDTSGGGGLQHS--GV 190
DB 122 EEQVLETESGAILVTSGLSPGVNVEKLTQLISAQKGRICITDSSGDALTALALGDI 181
OY 191 FLKASVRELRRECVCSEILTEPEQL-AAAHELIDRGRAEVVVVSLGSGALLATRHASHR 249
DB 182 ELVFNKELSLALVNRD-LIQPDVVRKAQELVSGKARVVSILGFGALGISSENCIQ 240
OY 250 FSSITPMITAVSGVAGDMVAALITVGLSRGWSLIRVPLGNAGAAMLITPTAACNRDV 309
DB 241 VVPPPVKSQSTVGAGDSMGVAMTKLQADASLEEMVRFVGAAGSAATLNOGTRLCSHDDT 300
OY 310 ERFP 313
DB 301 QKIT 304
XX
XX RESULT 10
XX AUJ34587
XX ID AUJ34587 standard; protein; 309 AA.
XX
XX AC AUJ34587;
XX
XX DT 14-FEB-2002 (first entry)
XX
XX DE E. coli cellular proliferation protein #168.
XX
XX KM Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.
XX
XX OS Escherichia coli.
XX
XX PN WO200170955-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 21-MAR-2001; 2001WO-US009180.
XX
XX PR 21-MAR-2000; 2000US-0191078P.
XX PR 23-MAY-2000; 2000US-0206848P.
XX PR 26-MAY-2000; 2000US-0207727P.
XX PR 23-OCT-2000; 2000US-0242578P.
XX PR 27-NOV-2000; 2000US-0253625P.
XX PR 22-DEC-2000; 2000US-0257931P.
XX PR 16-FEB-2001; 2001US-0269308P.
XX
XX PA (ELIT-) ELITRA PHARM INC.
XX
XX PI Haselbeck R, Ohlsen KU, Zykind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX DR WPI; 2001-611495/70.
XX DR N-PSDB; AAS52446.
XX
XX PT New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids.
XX
XX XX Example 3; SEQ ID NO 10180; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the genes,
XX their use in the discovery of novel antibiotics, the essential genes
XX themselves and the encoded proteins. The prokaryotes used are Escherichia
XX coli, Staphylococcus aureus, Salmomella typhi, Klebsiella pneumoniae,
XX Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
XX useful for the identification of potential new targets for antibiotic
XX development. The antisense nucleic acids can also be used to identify
XX proteins used in proliferation, to express these proteins, and to obtain
XX antibodies capable of binding to the expressed proteins. The proteins can
```



XX ADN18198;  
AC  
XX  
XX 02-DEC-2004 (first entry)  
DT  
XX  
XX Bacterial polypeptide #851.  
DE  
XX  
XX Recombinant DNA construct; transformed plant; improved plant property;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
KW pathogen tolerance; pest tolerance; plant disease resistance;  
KW cell cycle pathway modification; plant growth regulator;  
KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
KW bacterial polypeptide.  
XX  
XX Bacteria.  
OS  
XX  
XX US2003233675-A1.  
PN  
XX  
XX 18-DEC-2003.  
PD  
XX  
XX 20-FEB-2003; 2003US-00369493.  
PF  
XX  
XX 21-FEB-2002; 2002US-0360039P.  
PR  
XX  
XX  
XX (CAOY/) CAO Y.  
PA (HINK/) HINKLE G J.  
PA (SLAT/) SLATER S C.  
PA (CHEN/) CHEN X.  
PA (GOLD/) GOLDMAN B S.  
XX  
XX  
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
PI  
XX  
XX WPI: 2004-061375/06.  
DR  
XX  
XX New recombinant DNA construct comprising a promoter positioned to provide  
PT for expression of a polynucleotide encoding a polypeptide from a  
PT microbial source, useful for producing plants with improved properties.  
PT  
XX  
XX Claim 1; SEQ ID NO 851; 122pp; English.  
PS  
XX  
XX The invention relates to a recombinant DNA construct comprising a  
CC promoter functional in a plant cell, where the promoter is positioned to  
CC provide for expression of a polynucleotide encoding a polypeptide from a  
CC microbial source. The invention also relates to a transformed plant  
CC comprising the recombinant DNA construct and a method of producing a  
CC transformed plant having an improved property. The plant is a crop plant  
CC such as maize or soybean. The method of producing a transformed plant  
CC having an improved property comprises transforming a plant with the  
CC recombinant DNA construct and growing the transformed plant, where the  
CC polynucleotide or polypeptide is useful for improving plant properties.  
CC The recombinant DNA construct is useful for producing plants with  
CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
CC increased resistance to plant disease, better growth rate by modification  
CC of the cell cycle pathway with plant growth regulators, increased rate of  
CC homologous recombination, modified seed oil or protein yield and/or  
CC content, improved yield by modification of carbonate, nitrogen or  
CC phosphorus use and/or uptake, by modification of photosynthesis or by  
CC providing improved plant growth and development under at least one stress  
CC condition, improved lignin production or improved galactomannan  
CC production. This sequence represents a bacterial polypeptide used in the  
CC scope of the invention. Note: The sequence data for this patent did not  
CC form part of the printed specification but was obtained in electronic  
CC format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
XX  
XX Sequence 309 AA;

Db 3 RYITTLTLPSSLDASATITTPQIYIEGKLRCTAPFEEGGGGINARAIHLGGSTALFPAG 62

QY 73 GSTGSLMALIGDAGVPRVPVPIAASTBSFTVNSRTAKOYRFVLPGBSLTVAEOQCI 132

Db 63 GATGSHVLSLLADENVVPVAIVEAKOMTRQNHLHVHASGEQYRFVMPGALNDEFERQ-L 121

QY 133 DELRGAAASAFVVASGSLPGCVADYYQYRVADICRRSTPLILDTSSGGGLCHISS--CY 190

Db 122 EEOVLIEISGAILLVISGSLPGCVKLEKUTQLLSAOKOGIRCIWSSGGALSALALIGN 191

QY 191 FLTKASVELRECVGSELLTEPEOL-AAAHELIDRGAPVWVVSIGCSOGALATLRASHR 249

Db 182 ELVKNQKELSLAVNR-LTOPDDVRKAAQELVNSGKAKRVVVSIGPQALGVDSENCIO 240

QY 250 FSSIPMTAVSGVAGADANVAALITVGLSRGWSLIVSRIGNAAGAAMLTPGTAACNRDDV 309

Db 241 VVPPVVKQSITVGAQDSMVGMNTLTLAENASLEEMVRFVAAGSAATLNOGRLCSHDP 300

QY 310 ERFF 313

Db 301 QKIT 304

RESULT 13  
 ABU28039  
 ID ABU28039 standard; protein; 311 AA.  
 XX  
 AC ABU28039;  
 XX  
 XX 19-JUN-2003 (first entry)  
 DT  
 DE Protein encoded by Prokaryotic essential gene #13566.  
 DE  
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 KW  
 OS Enterobacter cloacae.  
 XX  
 PN WO200271183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;  
 PI Wall D, Trawick JD, Carr GT, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-023926/02.  
 DR N-PSDB; ACAA31909.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 55963; 1766pp; English.  
 XX  
 XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
CC  
XX  
SQ Sequence 311 AA;

Query Match 30.9%; Score 531.5; DB 6; Length 311;  
Best Local Similarity 41.1%; Pred. No. 2.2e-45;  
Matches 125; Conservative 56; Mismatches 118; Indels 5; Gaps 4;

QY 14 IITITNMPALDITTSVDVVRPTEKRCGAPRYPDGGGGINNARIYHVHVGSGSTALFPAGG 73  
DB 5 IYTLTSPSLDSATLTPQIYPEGKLRCSAPVFEPPGGGINNARAITHLGGKATAIFFPAG 64  
QY 74 STGSLMALGDAVGPFRVPIPIASTRESFTVNESRTAKOYRFVLPGBSLTVAEOQCLD 133  
DB 65 ATGHLVSLADEQVAETVAKWTRONLHVHVGSGSEQRRFVMPAKLSDNFFRQ-LE 123  
QY 134 ELRGAAASAAFFVVASGSLPGVADYQVRVADICRSTPLILDTSGGGLQH--ISSGVF 191  
DB 124 EKVLITISGALLVTSGLPFCVSTFKLTALIQAAQVRIRCTVDSGSLQALVPGNLB 183  
QY 192 LUKASVRELRECVSELTTEPEQL-AAAHELIDRGAEVYVVSLSGSGALLATRHASHRF 250  
DB 184 LVENQKELSLVNR-LTOPDDVRTAAOELVRTGKARRVVSLSGPGALAVDESSVQV 242  
QY 251 SSIPMTAVSGVAGDAMVAATVGLSRGWSLTKSVRLGNAGAAAMLTPGTAAACNRDVE 310  
DB 243 VPPPMKSQSTVAGADSMVGAWTLKLAQASLLEMTIRIGVAAAGSAATTINQTRLCSLADTQ 302  
QY 311 RFFE 314  
DB 303 KIVD 306

RESULT 14  
ABO63860  
ID ABO63860 standard; protein; 314 AA.  
XX  
AC ABO63860;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Klebsiella pneumoniae polypeptide seqid 103177.  
XX  
KM Recombinant expression vector; transcription regulatory element;  
XX Klebsiella pneumoniae proteoin; antibacterial; Vaccine.  
XX  
OS Klebsiella pneumoniae.  
XX  
PN US6610836-B1.  
XX  
PD 26-AUG-2003.  
XX

PF 27-JAN-2000; 2000US-00489039.  
XX  
PR 29-JAN-1999; 99US-0117747P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
PA  
XX Breton GL, Osborne M;  
PI  
XX WPI; 2003-895346/82.  
DR N-PSDB; ACH97411.  
XX  
XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
PT preparing a vaccine composition against Klebsiella pneumoniae.  
PT  
XX Disclosure; SEQ ID NO 10377; 932bp; English.  
XX  
XX The invention describes a new isolated nucleic acid encoding a Klebsiella  
XX pneumoniae polypeptide. Also described are: a recombinant expression  
XX vector comprising the nucleic acid, operably linked to a transcription  
XX regulatory element; and a cell comprising the recombinant expression  
XX vector. The nucleic acid is useful for preparing a vaccine composition  
XX against Klebsiella pneumoniae. This is the amino acid sequence of a  
XX Klebsiella pneumoniae polypeptide of the invention  
XX  
SQ Sequence 314 AA;

Query Match 30.3%; Score 520.5; DB 7; Length 314;  
Best Local Similarity 39.6%; Pred. No. 3e-44;  
Matches 122; Conservative 52; Mismatches 131; Indels 3; Gaps 2;

QY 9 EGRPRITITLTMNPAIDITTSVDVVRPTEKRCGAPRYPDGGGGINNARIYHVHVGSGSTAL 68  
DB 3 EGMTKIYTLTAPSLDSATQTPQIYPEGKLRCSAPVFEPPGGGGINNARAATFLGGKATAI 62  
QY 69 FPAAGSTGSLMALGDAVGPFRVPIPIASTRESFTVNESRTAKOYRFVLPGBSLTVAEQ 128  
DB 63 FVVGATGEHLAALLADEQVPVETVETRDWTRONLHVHVAASSEQRRFVMPALIDDEF 122  
QY 129 EOCIDELRGAAASAAFFVVASGSLPGVADYQVRVADICRSTPLILDTSGGGLQH-- 186  
DB 123 RR-LBEKVLITTEPSSLVTSGLPFGISVDNLMQVGNAGQGLRCITDSSGDLAALD 181  
QY 187 SSGVFLIKASVRELRECVSELTTEPEQLAAAHELIDRGAEVYVVSLSGSGALLATRHA 246  
DB 182 VGNIELVKNQKELSLVNRDLQPDVRLAAOSLTIQSGVRRVVSLSGPGALGVDAAG 241  
QY 247 SHRFSSIPMTAVSGVAGDAMVAATVGLSRGWSLTKSVRLGNAGAAAMLTPGTAAACNR 306  
DB 242 SVQVPPPMKSQSTVAGADSMVGAWTLRLAENASLEDMVRFVGAAGSAATTINQTRLCSSR 301  
QY 307 DVERFFE 314  
DB 302 ANQOKIYD 309

RESULT 15  
ADH88496  
ID ADH88496 standard; protein; 318 AA.  
XX  
AC ADH88496;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Enterococcus faecalis polypeptide #2976.  
XX  
KM Enterococcus faecalis infection; transcription regulatory element;  
XX antibacterial.  
XX  
OS Enterococcus faecalis.  
XX  
PN US6617156-B1.  
XX  
PD 09-SEP-2003.  
XX



**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

**This Page Blank (uspto)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 05:11:18 ; Search time 13.4872 Seconds  
(without alignments)  
2418.401 Million cell updates/sec

Title: US-10-617-038-22

Perfect score: 1720  
Sequence: 1 MTEPAAWDEGKPRITLTLMN.....TEVGDDQYVWHPIVNPASP 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1720	100.0	339	2 D70942	probable pfkB prot
2	534.5	31.1	310	2 AB0707	6-phosphofructokin
3	531.5	30.9	309	1 K1ECB	6-phosphofructokin
4	530.5	30.8	309	2 A85781	6-phosphofructokin
5	530.5	30.8	309	2 E90932	6-phosphofructokin
6	318.5	18.5	315	2 C75626	1-phosphofructokin
7	309.5	18.0	308	2 B97263	tagatose-6-phospha
8	307	17.8	310	2 E43258	tagatose-6-phospha
9	304	17.7	309	2 B96928	1-phosphofructokin
10	300.5	17.5	303	2 A69627	fructose-1-phospha
11	283	16.5	306	2 D89841	fructose-1-phospha
12	275	16.0	307	2 A11735	fructose-1-phospha
13	268	15.6	318	2 A40944	1-phosphofructokin
14	267.5	15.6	310	2 S04358	1-phosphofructokin
15	266.5	15.5	310	2 D90015	1-phosphofructokin
16	266	15.5	306	2 C83753	fructose-1-phospha
17	266	15.5	307	2 A11366	fructose-1-phospha
18	263.5	15.3	307	2 E70178	1-phosphofructokin
19	257	14.9	309	2 A98006	tagatose-6-phospha
20	255.5	14.9	310	2 C39778	tagatose-6-phospha
21	250	14.5	309	2 C95138	tagatose-6-phospha
22	243.5	14.2	317	2 C82450	1-phosphofructokin
23	242	14.1	303	2 B95101	1-phosphofructokin
24	239	13.9	303	2 C97969	1-phosphofructokin
25	227	13.2	316	1 A39407	1-phosphofructokin
26	226	13.1	304	2 AB0888	probable carbohydr
27	223.5	13.0	312	2 AE0783	1-phosphofructokin
28	222.5	12.9	312	2 F85855	fructose-1-phospha
29	222.5	12.9	312	2 B37245	1-phosphofructokin

30	222.5	12.9	312	2 D91011	fructose-1-phospha
31	222	12.9	314	2 A83200	1-phosphofructokin
32	218.5	12.7	312	2 AF0158	1-phosphofructokin
33	206.5	12.0	305	2 F86744	tagatose-6-phospha
34	198	11.5	313	2 A64069	1-phosphofructokin
35	195	11.3	383	2 D75288	carbohydrate kinase
36	187.5	10.9	319	2 H72327	sugar kinase, pfkB
37	174.5	10.1	310	2 AD1707	1-phosphofructokin
38	173.5	10.1	310	2 AG1336	1-phosphofructokin
39	170	9.9	483	2 F87700	rfae protein (lipo
40	153	8.9	313	2 AG3520	ribokinase (EC 2.7
41	152.5	8.9	299	2 E72311	ribokinase - Therm
42	151	8.8	303	2 B87367	ribokinase (import
43	147	8.5	308	2 A10001	ribokinase (EC 2.7
44	144	8.4	319	2 AC0942	probable carbohydr
45	142	8.3	311	2 E75385	carbohydrate kinase

ALIGNMENTS

RESULT 1

D70942  
probable pfkB protein - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004

C/Accession: D70942

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;

Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.;

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; PMID:98295987; PMID:9634230

A/Accession: D70942

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-339 <COL>

A/Cross-references: UNIPROT:O86352; UNIPARC:UPI00000D1173; GB:AL021899; GB:AL123456; NID

A/Experimental source: strain H37RV

C/Genetics:

A/Gene: pfkB

C/Superfamily: 6-phosphofructokinase 2

Query Match 100.0%; Score 1720; DB 2; Length 339;  
Best Local Similarity 100.0%; Pred. No. 7; 7e-117;  
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MTEPAAWDEGKPRITLTLMN	PALDITTSVDVVRPTKMRGAGPRYDPGGGGINVARIHV	60
DB	1	MTEPAAWDEGKPRITLTLMN	PALDITTSVDVVRPTKMRGAGPRYDPGGGGINVARIHV	60
QY	61	LGCGSTLFPAGSGTSLMAL	LGDAQVPRVPIPASTESTTVNSTRAKQYRFLPG	120
DB	61	LGCGSTLFPAGSGTSLMAL	LGDAQVPRVPIPASTESTTVNSTRAKQYRFLPG	120
QY	121	PSLTVAEQECDELGAASAF	VVAGSLPGVAADYQRYADICRSSTPLIDTSG	180
DB	121	PSLTVAEQECDELGAASAF	VVAGSLPGVAADYQRYADICRSSTPLIDTSG	180
QY	181	GGLQHTSSGVFLKASVRE	LCVSGSELTTEPQLAAHBLIDRGRAEVVVSLSGQAL	240
DB	181	GGLQHTSSGVFLKASVRE	LCVSGSELTTEPQLAAHBLIDRGRAEVVVSLSGQAL	240
QY	241	LATRHASHRRSSIPMTN	AVSGAGDMVAITVGLSRGWSLISVPLGNAAGAMLLTPG	300
DB	241	LATRHASHRRSSIPMTN	AVSGAGDMVAITVGLSRGWSLISVPLGNAAGAMLLTPG	300
QY	301	TAACNRDVERFPFLAAPE	TEVGDDQYVWHPIVNPASP 339	
DB	301	TAACNRDVERFPFLAAPE	TEVGDDQYVWHPIVNPASP 339	

## RESULT 2

AB0707  
6-phosphofructokinase isozyme [imported] - *Salmonella enterica* subsp. *enterica* serovar Typh  
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typh  
A:Note: this species has also been called *Salmonella typhi*  
C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AB0707  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AB0707  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-310 <PAR>  
A:Cross-references: UNIPARC:UP10000059DC1; GB:AL513382; PIDN:CAD02027.1; PID:G16502864;  
C:Genetics:  
C:Gene: STY1785  
C:Superfamily: 6-phosphofructokinase 2

Query Match 31.1%; Score 534.5; DB 2; Length 310;  
Best Local Similarity 41.4%; Pred. No. 2.5e-31;  
Matches 126; Conservative 54; Mismatches 119; Indels 5; Gaps 4;  
Qy 13 RITLTMPALDITTSVDVVRPTEKRCGAPRYPDGGGGINVARIYVHVGCGCTALFPAG 72  
Db 3 RITLTMPALDITTSVDVVRPTEKRCGAPRYPDGGGGINVARIYVHVGCGCTALFPAG 62  
Qy 73 GSTGSLMALLGAGVPRVPIAASSTRESFTNESRTAKOYRVLPGPSITVAEQOCL 132  
Db 63 GATGEHLVSLADENPVATVEAKDMTRQNLHVHVASGEQYRFVWGATLIDDEFRO-L 121  
Qy 133 DELRGAASAFAVVASGSLPPGVADYQYRVADI CRSSSTPLIDTSGGGLQHISS--GV 190  
Db 122 EBOVLEIESGALIVISGSLPPGVAVKLTQLSAQAQKGRICITDSGDALTALAGDI 181  
Qy 191 FLTKASVREIRECVSEELTEPEQL-AAAHELIDRGRAVYVVSLSQGLALTRASHR 249  
Db 182 ELVKPNKELISALVNRD-LTQPDVVRKAQGLVSGKARRVYVSLGQGLGIDSENCIQ 240  
Qy 250 FSSIPMTAVSGVAGDAMVAITVGLSRGWSLTKSVRLGNAAAGAMLLTPGTAAACNRDV 309  
Db 241 VVPPVKSQSTVGAGDSMGVAMTKLAQDASLEEMVRFVVAAGSAATLNOGTRLCSHDT 300  
Qy 310 ERF 313  
Db 301 QKIV 304

## RESULT 3

KIECFB  
6-phosphofructokinase (EC 2.7.1.11) isozyme 2 - *Escherichia coli* (strain K-12)  
N:Alternate names: phosphofructokinase 1 isozyme 2; phosphohexokinase isozyme 2  
C:Species: *Escherichia coli*  
C>Date: 28-Dec-1987 #sequence\_revision 17-Oct-1997 #text\_change 09-Jul-2004  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Berna, N.T.; Burland, V.; Riley, M.; Co  
A.: Rose, D.J.; Mau, B.; Sano, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of *Escherichia coli* K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: G64931  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-309 <BIAT>  
A:Cross-references: UNIPROT:P06999; UNIPARC:UP1000012DB3D; GB:AE000267; GB:U00096; NID:G  
R:Delid, F.  
Gene 28, 337-342, 1984  
A:Title: Nucleotide sequence of gene pfk encoding the minor phosphofructokinase of *Esch*

A:Reference number: A91513; MUID:84262485; PMID:6235149

A:Accession: A24950  
A:Molecule type: DNA  
A:Residues: 1-25; 'ENCAVPHRCSN', 37-40, 42-154, 'LRKNK', 160, 'SAASSTVL', 169, 'QC', 172-244, 'AL',  
A:Cross-references: UNIPARC:UP1000016F3C7; GB:K02500; NID:G147151; PIDN:AAA24321.1; PID:G  
A:Experimental source: strain K12  
R:Delid, F.  
J. Mol. Biol. 168, 285-305, 1983  
A:Title: Molecular cloning of the gene for phosphofructokinase-2 of *Escherichia coli* and  
A:Reference number: I56378; MUID:83294514; PMID:6310120  
A:Accession: I56378  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-25; 'ENCAVPHRCSN', 37-38 <RBS>  
A:Cross-references: UNIPARC:UP1000016F3C6; GB:K00128; NID:G147149; PIDN:AAA24320.1; PID:G  
C:Comment: This enzyme is not to be confused with 6-phosphofructo-2-kinase (EC 2.7.1.105)  
C:Genetics:  
A:Gene: pfkB  
A:Map position: 38 min  
A:Map complex: homotrimer  
C:Function:  
A:Description: catalyzes phosphorylation of fructose 6-phosphate to fructose 1,6-bisphos  
A:Pathway: glycolysis  
C:Superfamily: 6-phosphofructokinase 2  
C:Keywords: glycolysis; phosphotransferase

Query Match 30.9%; Score 531.5; DB 1; Length 309;  
Best Local Similarity 41.1%; Pred. No. 4e-31;  
Matches 125; Conservative 55; Mismatches 119; Indels 5; Gaps 4;  
Qy 13 RITLTMPALDITTSVDVVRPTEKRCGAPRYPDGGGGINVARIYVHVGCGCTALFPAG 72  
Db 3 RITLTMPALDITTSVDVVRPTEKRCGAPRYPDGGGGINVARIYVHVGCGCTALFPAG 62  
Qy 73 GSTGSLMALLGAGVPRVPIAASSTRESFTNESRTAKOYRVLPGPSITVAEQOCL 132  
Db 63 GATGEHLVSLADENPVATVEAKDMTRQNLHVHVASGEQYRFVWGATLIDDEFRO-L 121  
Qy 133 DELRGAASAFAVVASGSLPPGVADYQYRVADI CRSSSTPLIDTSGGGLQHISS--GV 190  
Db 122 EBOVLEIESGALIVISGSLPPGVAVKLTQLSAQAQKGRICITDSGDALTALAGDI 181  
Qy 191 FLTKASVREIRECVSEELTEPEQL-AAAHELIDRGRAVYVVSLSQGLALTRASHR 249  
Db 182 ELVKPNKELISALVNRD-LTQPDVVRKAQGLVSGKARRVYVSLGQGLGIDSENCIQ 240  
Qy 250 FSSIPMTAVSGVAGDAMVAITVGLSRGWSLTKSVRLGNAAAGAMLLTPGTAAACNRDV 309  
Db 241 VVPPVKSQSTVGAGDSMGVAMTKLAQDASLEEMVRFVVAAGSAATLNOGTRLCSHDT 300  
Qy 310 ERF 313  
Db 301 QKIV 304

## RESULT 4

A85781  
6-phosphofructokinase II, suppressor of pfkA [imported] - *Escherichia coli* (strain O157:f  
C:Species: *Escherichia coli*  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
R:Perma, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,  
iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: A85781  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-309 <STO>  
A:Cross-references: UNIPROT:Q8XE22; UNIPARC:UP100000028F; GB:AE005174; NID:G12515736; P  
A:Experimental source: strain O157:H7, substrain EDL533  
C:Genetics:

A:Gene: pfkB  
C:Superfamily: 6-phosphofructokinase 2

Query Match	30.8%;	Score 530.5;	DB 2;	Length 309;
Best Local Similarity	41.1%;	Pred. No. 4.8e-31;		
Matches 125;	Conservative 54;	Mismatches 120;	Indels 5;	Gaps 4;

```

Oy 13 IITLTMPALDITTSVQVVRPTECMRGCAVPYDGGGINVARVHVLGSCFALPAG 72
Db 3 RIYTLTAPSDSAITTPQIYIEGKURCTAIVFEFGGGGINVAIAIAHGGSAIRIIFAG 62
Oy 73 GSTGSLMALIGDAGVFPFRVPIAIASTSESTVNESRTAKOYRFVLPBPSLTVAEOECL 132
Db 63 GATGHEHLTLADENVVPATVEAKDMTQONHIVEASGEQYRFMPAPALNEBFRO-L 121
Oy 133 DELRGAAASAFAVVASGSLPGVADYVORVADICRRSSTPLIDTSGGLHSHS--GV 190
Db 122 EEOYVEISGAINIVISGSLPGVVKLEKLTQLISAAQNGIRNCIIDSGEALSAALAIQNI 161
Oy 191 FLTKASVRELBECVGSSELLTEPEOL-AAAHLEIDRGRAEVVVVISGOGALLATRIASHR 249
Db 182 ELVKNQOELSAVURE-LTQPDVYRKAQOETIVNSGKRAVVSISGPGAGVDSENCIO 240
Oy 250 FSSIEMTAVSGAGADAAVAAITVGLSRGWSLKSRLGNAGAAMLLTPTAACNREDV 309
Db 241 VVPPPVKSQSTVAGAGDSVAGMTLKLAMASILEEYVRGVAAGSAATLNGSTRICSHDT 300
Oy 310 ERFF 313
Db 301 QKIT 304

```

## RESULT 5

E90932  
6-phosphofructokinase II [imported] - *Escherichia coli* (strain O157:H7, substrain RIMD 0001/01)  
C:Species: *Escherichia coli*  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C:Accession: E90932  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.-G. gasaara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic islands  
A:Reference number: A59629; MUID:21156231; PMID:11258796  
A:Accession: E90932  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-309 <HAY>  
A:Cross-references: UNIPROT:O8XE22; UNIPARC:UPI00000D02BF; GB:BA000007; PIDD:BA835852.1  
A:Experimental source: strain O157:H7, substrain RIMD 050952  
C:GeneticDB:  
A:Gene: EC62429  
A:Superfamily: 6-phosphofructokinase 2

	Query Match	30.8%	Score 530.5	DB 2	Length 309	
	Best Local Similarity	41.1%	Pred. No. 4.8e-31			
	Matches	125	Conservative	54	Mismatches	120
					Indels	5
					Gaps	4
OY	RIITLTMPALDITTSVDVVRPTEKRCGABRYDPGGGINVARIVHYLGGCSTALPAPAG	72				
Dd	RIITLTLPALSDATITPQIYPBGKLRCTALPFVEPGGGINVARIAHLGGSATAIIPAG	62				
OY	3 GSTGSLMALLDGAVPRFVIPPIASTRESFTVAESRPAKQYRFLLPPSPILTVAEQECL	132				
Dd	63 GATGHTLVSLADENVPAATVEAKDMTRONLHVHEASGEQYRFPMFPALANEDFR-L	121				
OY	133 DELRGAAASAFFVVASGSILPGVADYYQRVADICRRSSPTLIDTSGGGLHITS--GV	190				
Dd	122 EEOVLETSIGAILVTSSGLPGCVTKLEKLTOLSAQNQCINCIIIDSSSEALSALAIGNI	181				
OY	191 FLUKASVRELRECVSSELLTEPEQL-AAAHHELIDGRALVVVVYSIGSGALLATRHASHR	249				
Dd	182 ELVKNQNELSALVNR-LTPQDDVTKAAQETIVNSGKRKRVVSLGPFGALGVDSENCIO	240				

OY	250	FSSTPMTAVSGVAGDAMVAALTITVLGSRGWSLTKSVRLGNNAAGAAMLTTPCTAACNRDPV	309
		:           :           :	
Db	241	VPEPVKSSGTIVAGSDMYGATMLKLDAENASLEENRYRGVAAGSAAATLTNGSTRCSHDDT	300
		:           :           :	
OY	310	ERFP	313
	:	:	:
Db	301	QRTY	304

## RESULT 6

C75626  
 1-phosphofructokinase - *Deinococcus radiodurans* (strain R1)  
 C.Species: *Deinococcus radiodurans*  
 C.Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C.Accession: C75626  
 R.White, O.; Eslen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; I  
 R.M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Mal  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A.Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.  
 A.Reference number: R75250; MUID:20036896; PMID:10567266  
 A.Accession: C75626  
 A.Status: preliminary  
 A.Molecule type: DNA  
 A.Residues: 1-315 <WH1>  
 A.Cross-references: UNIPROT:Q9RZP6; UNIPARC:UPI0000003B4E; GB:AE001826; NID:G6460827; PII  
 A.Experimental source: strain R1  
 C.Genetics:  
 A.Gene: DRB0074  
 A.Map position: megaplasmid  
 A.Genome: plasmid  
 A.Note: plasmid MP1  
 C.Superfamily: 6-phosphofructokinase 2

[illegible]

## RESULT 7

B97263  
tagatose-6-phosphate kinase [imported] - Clostridium acetobutylicum  
C.Species: Clostridium acetobutylicum  
C.Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C.Accession: B97263  
R.NOLLING, J.; Breton, G.; Omelchenko, M.V.; Markova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
Article: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum



A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallet  
leach, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardiniois,  
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y. M.; Ogawa, K.; Ogiwara, A.; Oudgaa, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadleir, Y.; Sato, T.; Scanlon,  
A;Authors: Schleich, S.; Schroeter, R.; Seftone, P.; Sekiguchi, J.; Sekowska, A.; Serot  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A;Authors: Yoshikawa, H.F.; Zumein, E.; Yoshikawa, H.; Danchin, A.  
A;Title: The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*.  
A;Reference number: A69580; MUID:96044033; PMID:9384377  
A;Accession: A69627  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-303 <KUN>  
A;Cross-references: UNIPROT:O31714; UNIPARC:UPI000006032D; GB:I299111; GB:AL009126; NID:G  
A;Experimental source: strain 168  
C;Genetics:  
A;Gene: frub  
C;Superfamily: 6-phosphofructokinase 2

Query Match 17.5%; Score 300.5; DB 2; Length 303;  
Best Local Similarity 29.4%; Pred. No. 1.8e-14;  
Matches 98; Conservative 59; Mismatches 115; Indels 61; Gaps 14;  
Qy 14 IITLNNPALDITTSVDVVRPTEKMRGCA---PRYD---PGGGGINVARIY---HVLGGC 64  
Db 2 IYTTNLP-----SVDYIVVEDPTVGLRRSYDTYRPGKGINSRLLKRRHV---A 52  
Qy 65 STALFPAGSGTSLMLLGDAGVPRVPIAASSTRESFTVN-ESRTAKQYRFLVPGESL 123  
Db 53 SKALFGVGGFTGERTIKTFLRENIETARSVKGDR-----INVKLKTEDEINGQGPFI 108  
Qy 124 T-----VAEBOCCIDELRGAAASAFVYASGSLPFGVADYQYQVADICRRSSTPLILD 177  
Db 109 SDEDFKAFLEQFQSQE-----GDIYVLAGSIPSSLPHTYKELIACCKQONARVLLD 161  
Qy 178 TSGGG-LQHSISGVFLKASVRELRECVSESLTEPEOLAANEHLIDRGRAVYVVSLSGS 236  
Db 162 ISGGLKATKEMKPFLEKPNHELGEMGTATITSTEVAVPFGKLVLEG-AERIVYSAG 220  
Qy 237 QGALLATRHASHRPSIPW-TAVSGVAGADMAAIVTGLSRGSLIKSVRLGNAAAGAM 295  
Db 221 DGAALLFTBEAVY-FANVVKGLVNSVAGDSVAGFLAGISKOJPLEBAPRLGVTSSGA- 278  
Qy 296 LITPETAACNRDVERPFLAABEPVCGQDQYV 328  
Db 279 -----TAFSEBELGTBEFV 291

RESULT 11  
D89841  
fructose-1-phosphate kinase [imported] - *Staphylococcus aureus* (strain N315)  
C;Species: *Staphylococcus aureus*  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C;Accession: D89841  
R;Kiroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratake, K.  
Lancet 357, 1225-1240, 2001  
A;Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
A;Reference number: A89758; MUID:21311952; PMID:11418146  
A;Accession: D89841  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-306 <KUN>  
A;Cross-references: UNIPROT:O9KWK1; UNIPARC:UPI00000CA92D; GB:BA000018; PID:g13700590; F  
A;Experimental source: strain N315  
C;Genetics:  
A;Gene: frub  
C;Superfamily: 6-phosphofructokinase 2

Query Match 16.5%; Score 283; DB 2; Length 306;  
Best Local Similarity 26.6%; Pred. No. 3.4e-13;  
Matches 81; Conservative 66; Mismatches 139; Indels 18; Gaps 7;  
Qy 14 IITLNNPALDITTSVDVVRPTEKMRGCAPRYPGGGGINVARIYVJGCGSTALFPAGG 73  
Db 2 IYTTNLPSSIDYVITNDPFIKIDGNRATATYKRAAGGINSRVKLTLVDSTRLGACG 61  
Qy 74 STGSLMALLDGAGVPRVPIAASSTRESFTVN-ESRTAKQYRFLVPGESLVAEOQCL 132  
Db 62 FPGKFIDTLNNSAIGNFIEVEDER-----INVKLKTEGETEINAPGHITSTQEQQL 117  
Qy 133 DELRGAAASAFVYASGSLPFGVADYQYQVADICRRSSTPLILDTSGGGLQHSISV-- 190  
Db 118 QQIKRN-TTSSDVIYVAGSVSSISPDYKQIAQITAGTGAKLVDAB-----KEIAESVLP 172  
Qy 191 ---FLKASVRELRECVSESLTEPEOLAANEHLIDRGRAVYVVSLSGSQALATRHAS 247  
Db 173 FHPFLFKNDLEFWMNTTVNSDPTVYIKGRLLVDKG-AQSVYVLSGGCGATYIDKEIS 231  
Qy 248 HRFSSIPMTAVSGVAGADMAVAAITVGLSRGSLIKSVRLGNAAAMLLTPGTACNRD 307  
Db 232 IKAVNPGKVVNTVYSGSDSTVAGVAGIAGLTERKAFQQAACGTATFDEDLA--TRD 289  
Qy 308 DVER 311  
Db 290 AIEK 293

RESULT 12  
A11735  
fructose-1-phosphate kinase [imported] - *Listeria innocua* (strain Clp11262)  
C;Species: *Listeria innocua*  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C;Accession: A11735  
R;Glaeser, P.; Franjeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgey, O.; Entian, K.D.; Fein, H.  
Science 294, 849-852, 2001  
A;Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat  
ok, C.; Schluter, T.; Stimes, N.; Tietze, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A;Title: Comparative genomics of *Listeria species*.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: A11735  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-307 <GLA>  
A;Cross-references: UNIPROT:Q928V2; UNIPARC:UPI00000CC878; GB:AL592022; PIDN:CA097657.1;  
A;Experimental source: strain Clp11262  
C;Genetics:  
A;Gene: frub  
C;Superfamily: 6-phosphofructokinase 2

Query Match 16.0%; Score 275; DB 2; Length 307;  
Best Local Similarity 27.3%; Pred. No. 1.3e-12;  
Matches 77; Conservative 58; Mismatches 141; Indels 6; Gaps 4;  
Qy 14 IITLNNPALDITTSVDVVRPTEKMRGCAPRYPGGGGINVARIYVJGCGSTALFPAGG 73  
Db 2 IYTTNLPSSIDYVITNDPFIKIDGNRATATYKRAAGGINSRVKLTLVDSTRLGACG 61  
Qy 74 STGSLMALLDGAGVPRVPIAASSTRESFTVNESRTAKQYRFLVPGESLVAEOQCLD 133  
Db 62 FTGGFIDWLKNEIGIQGFYIVDDPRINIKLKHGEETELNG--GPAISKEIKFEL- 117  
Qy 134 ELRGAAASAFVYASGSLPFGVADYQYQVADICRRSSTPLILDTSGGG-LQHSISGVFL 192  
Db 118 EVMDDKTVAGDIVILISGVPSPSLGNDPFYDKTIOICKEKQAFMIDTTCQELIDALPNRPIL 177  
Qy 193 LKASVRELRECVSESLTEPEOLAANEHLIDRGRAVYVVSLSGSQALATRHASHRFS 252  
Db 178 IKPNHHELELFVGLKDSVEDLIPYKKCLELG-AQHIVVSMAGDALTLTGEDVYPADA 236



Job time : 14.4872 secs

---



**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 04:56:43 ; Search time 86.7035 Seconds  
(without alignments)  
2758.529 Million cell updates/sec

Title: US-10-617-038-22  
Perfect score: 1720  
Sequence: 1 MTEPAAWDEKPRILITLTNM.....TEVGQDQYVHPIVPEASP 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues  
Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_05.80.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1720	100.0	339	Q7TZ43_MYCBO	Q7TZ43 mycobacteri
2	1720	100.0	339	Q86352_MYCTU	Q86352 mycobacteri
3	749	43.5	308	Q4J5V5_AZOVU	Q4J5V5 azobacter
4	667	38.8	347	Q4NK08_9M1CC	Q4NK08 mycobacte
5	648	37.7	312	Q6KCU0_RHOPA	Q6KCU0 rhodospseudo
6	603.5	35.1	313	Q631V9_BURPS	Q631V9 burkholderi
7	597.5	34.7	313	Q62EC2_BURMA	Q62EC2 burkholderi
8	574	33.4	295	Q89189_BRATA	Q89189 bradyrhizob
9	539.5	31.4	310	Q57PV8_SALCH	Q57PV8 salmonella
10	539.5	31.4	310	Q8ZPT5_SALTY	Q8ZPT5 salmonella
11	534.5	31.1	310	Q8Z6H7_SALPA	Q8Z6H7 salmonella
12	534.5	30.9	309	Q5PH98_SALPA	Q5PH98 salmonella
13	531.5	30.9	309	K6PFI2_ECOLI	P06999 escherichia
14	530.5	30.8	309	Q83140_SHIFL	Q83140 shigella fl
15	530.5	30.8	309	Q8XEE2_ECO57	Q8XEE2 escherichia
16	524.5	30.5	310	Q8RH20_ECOL6	Q8RH20 escherichia
17	355	20.6	317	Q8ZH57_STRAW	Q8ZH57 streptomyce
18	355	20.6	326	Q4NDB8_9M1CC	Q4NDB8 arthrobacte
19	354	20.6	307	Q73A75_BACCI	Q73A75 bacillus ce
20	354	20.6	310	Q9KXV8_STRCO	Q9KXV8 streptomyce
21	349	20.3	315	Q9KXV8_STRCO	Q9KXV8 streptomyce
22	348	20.2	307	Q63CY0_BACIZ	Q63CY0 bacillus ce
23	343.5	20.0	323	Q67Q12_STWTH	Q67Q12 symbiodacte
24	334	19.4	332	Q4NKX2_9M1CC	Q4NKX2 arthrobacte
25	330	19.2	309	Q8ZGF5_STRAW	Q8ZGF5 streptomyce
26	328	19.1	303	Q8ES06_OCEIH	Q8ES06 oceanobacil
27	327.5	19.0	326	Q4H6K1_9BEO	Q4H6K1 enterococcu
28	322	18.7	313	Q833W9_ENTFA	Q833W9 enterococcu
29	320	18.6	321	Q67RB4_STWTH	Q67RB4 symbiodacte
30	318.5	18.5	315	Q9ZRP6_DEIRA	Q9ZRP6 deinococcus
31	316	18.4	292	Q6AD91_LEIYX	Q6AD91 leifsonia x

32	314	18.3	303	2	Q73B5_BACCI	Q73B5 bacillus ce
33	312.5	18.2	315	2	Q6AG92_LEIYX	Q6AG92 leifsonia x
34	312	18.1	303	2	Q6HF93_BACHK	Q6HF93 bacillus ch
35	312	18.1	303	2	Q636V8_BACIZ	Q636V8 bacillus ce
36	311	18.1	303	2	Q81WV7_BACAN	Q81WV7 bacillus an
37	309.5	18.0	308	2	Q97F04_CLOAB	Q97F04 clostridium
38	309.5	18.0	316	2	Q5YV79_NOCRA	Q5YV79 nocardia fa
39	308	17.9	303	2	Q5K1W2_GEOXA	Q5K1W2 geobacillus
40	307	17.8	310	1	IACC_STRMU	P26421 streptococc
41	305	17.7	303	2	Q4MJY4_BACCE	Q4MJY4 bacillus ce
42	304	17.7	309	2	Q97MG6_CLOAB	Q97MG6 clostridium
43	304	17.7	315	2	Q88PQ4_PSEPK	Q88PQ4 pseudomonas
44	303	17.6	324	2	Q8PFP0_CORER	Q8PFP0 corynebacte
45	300.5	17.5	300	2	Q6Y9B0_BORHE	Q6Y9B0 borrelia he

## ALIGNMENTS

## RESULT 1

ID	Q7TZ43_MYCBO	PRELIMINARY;	PRT;	339 AA.
AC	Q7TZ43:			
DT	01-OCT-2003 (TREMBLrel. 25, Created)			
DT	01-OCT-2003 (TREMBLrel. 25, Last sequence update)			
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)			
DE	Probable phosphofructokinase pfkb (PHOSPHOKOKINASE)			
DE	(EC 2.7.1.-).			
GN	Name=pfkb; OrderedlocusNames=Mb2054c;			
OS	Mycobacterium bovis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium;			
OC	Mycobacterium tuberculosis complex.			
OX	NCBI_TaxID=1765;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE.			
RC	STRAIN=AF2122/97;			
RX	MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;			
RA	Garnier T., Sigmeleer K., Camus J.-C., Medina N., Mansoor H.,			
RA	Pryor M., Duthoy S., Gordin S., Iacox C., Monsemp C., Simon S.,			
RA	Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,			
RA	Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;			
FT	"The complete genome sequence of Mycobacterium bovis."			
RL	Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).			
DR	EMBL; BX248341; CAD96907.1; -; Genomic DNA.			
DR	GO; GO:0016301; F:kinase activity; IEA.			
DR	GO; GO:0016740; F:transferase activity; IEA.			
DR	InterPro; IPR002173; PFKB.			
DR	InterPro; IPR01611; PFKB_region.			
DR	Pfam; PF00294; PFKB; 1.			
DR	PROSITE; PS00583; PFKB_KINASES_1; 1.			
KW	Complete proteome; Kinase; Transferase.			
SO	SEQUENCE 339 AA; 35401 MW; D7735363FF73DADD CRC64;			

Query Match 100.0%; Score 1720; DB 2; Length 339;  
Best Local Similarity 100.0%; Pred. No. 2, 8e-109;  
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MTEPAAWDEKPRILITLTNMPALDITTSVDVVRTEKRCGAPRYDGGGINNARIYHV	60
DB	1	MTEPAAWDEKPRILITLTNMPALDITTSVDVVRTEKRCGAPRYDGGGINNARIYHV	60
QY	61	LGGSSTLFPAGSGTSLMALGADAGVPPRVIPRIASTRESFTVNSRTAKOYRFVLP	120
DB	61	LGGSSTLFPAGSGTSLMALGADAGVPPRVIPRIASTRESFTVNSRTAKOYRFVLP	120
QY	121	PSLTVABOEOGLDELRGAAASAFVVASGSLPPGVADYYQKVADICRSSTPLIDTSG	180
DB	121	PSLTVABOEOGLDELRGAAASAFVVASGSLPPGVADYYQKVADICRSSTPLIDTSG	180
QY	181	GGLHIISSGVPLLAASVRELRCEVSELTPEPELAAHLEIDRGAEVVVVVSGSGAL	240
DB	181	GGLHIISSGVPLLAASVRELRCEVSELTPEPELAAHLEIDRGAEVVVVVSGSGAL	240

Oy	241	LATRHASHRFSSTIPMTAVSGVAGADMAVAITMGSRGSLIKSVLGNAAAGAMLLTGG	300
Dd	241	LATRASHRFSSTIPMTAVSGVAGADMAVAITVGSRGSLIKSVLGNAAAGAMLLTGG	300
Oy	301	TACACRRDVERFEFFELAEPTVEGODQYVMHPINPEASP	339
Dd	301	TAACRRDVERFEFFELAEPTVEGODQYVMHPINPEASP	339
<hr/>			
RESULT 2	ID	086352 MYCTU PRELIMINARY;	PRT; 339 AA.
NC	086352	07D7L4;	
Dt	01-NOV-1998	(TREMBLrel. 08, Created)	
Dt	01-NOV-1998	(TREMBLrel. 08, Last sequence update)	
Dt	13-SEP-2005	(TREMBLrel. 31, Last annotation update)	
DE	Probable phosphofructokinase PFKB (PHOSPHOHEXOKINASE) (EC 2.7.1.-)		
DE	(Carbohydrate kinase, PFKB family).		
GN	Name=pfkb; OrderedLocustNames=M12008, Rv2029c;		
OS	Mycobacterium tuberculosis.		
OC	Bacteria; Actinobacteria; Actinomycetales;		
OC	Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;		
OC	Mycobacterium tuberculosis complex.		
OX	NCBI_Taxid=1773;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RC	STRAIN=H37RV;		
RX	MEDLINE=98295987,	PubMed=9634230,	DOI=10.1038/31159.
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Chutcher C.M.,		
RA	Harris D.E., Gordon S.V., Eigler K., Gas S., Barry C.E. Jr,		
RA	Tekela F., Badcock K., Basham D., Brown D., Chillingworth T.,		
RA	Connor R., Davies R.M., Devlin K., Fellwell T., Gentles S., Hamlin N.,		
RA	Holroyd S., Hornsby T., Jagels K., Krogh A., McLean U., Moule S.,		
RA	Murphy L.D., Oliver J., Osborne J., Quail M.A., Rajadream M.A.,		
RA	Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,		
RA	Sultson J.E., Taylor K., Whitehead S., Barrell B.G.,		
RT	Dispersing the biology of Mycobacterium tuberculosis from the		
RL	complete genome sequence."		
RM	Nature	393:537-544 (1998).	
RP	[2]		
RC	NUCLEOTIDE SEQUENCE.		
RX	STRAIN=CDC 1551 / Oshkosh;		
RX	MEDLINE=22206494;	PubMed=12218036;	
RX	DOI=10.1128/JB.184.19.5479-5490.2002;		
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,		
RA	Petersen J.D., Deboy R.T., Dodson R.J., Gilm M.L., Holt D.H.,		
RA	Hickey E.K., Kolonay J.F., Nelson W.C., Umeyam L.A., Ermolaeva M.D.,		
RA	Salzberg S.L., Delcher A., Utterback T.R., Weidman J.F., Khouri H.M.,		
RA	Gill Y., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,		
RA	Fraser C.M.;		
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and		
RL	J. Laboratory strains."		
DR	EMLB; BX842578;	CAAL1243.1;	-; Genomic DNA.
DR	EMBL; AE000516;	AAK46367.1;	-; Genomic DNA.
DR	PIR; D70942;	D70942.	
DR	TIGR; MT2088;	--	
DR	Tuberculin; Rv2029c;	--	
DR	GO; GO:0018301;	F:kinase activity;	IEA.
DR	InterPro; IPR002173;	PFKB.	
DR	InterPro; IPR011611;	PFKB_region.	
DR	Pfam; PF00294;	PEKB; 1.	
KW	PROSITE; PS00583;	PFKB_KINASES_1;	1.
SO	Complete proteome;	kinase; Transferase.	
SEQUENCE	339 AA;	35401 MM;	D773E353FF73DADD CRC64;
<hr/>			
Query Match	100.0%;	Score 1720;	DB 2; Length 339;
Best Local Similarity	100.0%;	Pred. No. 2.8e-109;	
Matches 339;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0
Oy	1	WTBPAMDGKRRIITLTMPALDITTSVDVVRPTEKRGCGAPRDPGGGGINVAITYV	60

```
Db      1 MTEPAMMBGKRRITLTMTNPLADITTSVDVVRPTENMKCAPRYDGGGGINVAIVHY 60
Qy      61 LGGCSTALFPAGSGTSLMALLGAGVPFRVIPAASTRFSPTVNESRTAKQYRFVLPG 120
Db      61 LGGCSTALFPAGSGTSLMALLGAGVPFRVIPAASTRFSPTVNESRTAKQYRFVLPG 120
Qy      121 PSLTYAEBOQCIDEIRGAASAAPFYVASGLSPGVAAIDYYORVADI CRSSSTPLLIDTSG 180
Db      121 PSLTYAEBOQCIDEIRGAASAAPFYVASGLSPGVAAIDYYORVADI CRSSSTPLLIDTSG 180
Qy      181 GGLHGISSCVFLLKXSVRELRCSVESELLTEPEQLAAAHELLDRGRAEVVVVSLSGQAL 240
Db      181 GGLHGISSCVFLLKXSVRELRCSVESELLTEPEQLAAAHELLDRGRAEVVVVSLSGQAL 240
Qy      241 LATRHASHRFFSSI PMTVAVSGVAGACDAMVAATVGLSRGSLIKSVRLGNAGAAMLTPG 300
Db      241 LATRHASHRFFSSIPMTAVS VGVAGDAMVAATVGLSRGSLIKSVRLGNAGAAMLTPG 300
Qy      301 TAACNRDDVERFPEELAEPTVEIGDOOYVNHPIYNPEASP 339
Db      301 TAACNRDDVERFPEELAEPTVEIGDOOYVNHPIYNPEASP 339

RESULT 3
Q4J5V5_AZOVI PRELIMINARY; PRT; 308 AA.
ID Q4J5V5_AZOVI PRELIMINARY; PRT; 308 AA.
AC Q4J5V5;
DT 13-SEP-2005 (TRMBLrel. 31, Created)
DT 13-SEP-2005 (TRMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TRMBLrel. 31, Last annotation update)
DE Carbohydrate kinase, FfkB.
GN ORFNames=AvINDRAFT_5718;
OS Azotobacter vinelandii AVOP.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
NC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxId=322710;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AVOP;
RA US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Iserant S., Pitluck S., Richardson P.;
RA "Sequencing of the draft genome assembly of Azotobacter vinelandii
RA AVOP.";
RT Submitted (JUN-2005) to the EMBL/genbank/DBJ databases.
RL [2]
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AVOP;
RA US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RA "Annotation of the draft genome assembly of Azotobacter vinelandii
RA AVOP.";
RT Submitted (JUN-2005) to the EMBL/genbank/DBJ databases.
RL [3]
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AVOP;
RA DOE Joint Genome Institute;
RA Submitted (FEB-2003) to the EMBL/genbank/DBJ databases.
RL [4]
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AVOP;
RA US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Iserant S., Pitluck S., Richardson P.;
RA Submitted (JUN-2005) to the EMBL/genbank/DBJ databases.
RL -1- CAUTION: The sequence shown here is derived from an
CC EMBL/genbank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AAU03000001; EAM07036.1; -; Genomic_DNA.
SQ KINASE.
EMBL; AAU03000001; EAM07036.1; -; Genomic_DNA.
SQ SEQUENCE 308 AA; 32503 MW; 035FD99DBC74190D CRC64;
```



Qy	193	IKASVETLRECVSELTTEBEQULAAHEIDRGRAVWVSLGSGALLARHSHSFSS	252
		1	2
Db	184	IKPNLRMOELARPJGCEKEWMDARAHITAGQVEYVALSTGHGALLVTKDRLSPA	243
		1	2
Qy	253	IPMTAVSGVAGDPAVTAATTVGLSRGWSLIKSVIRGNAGAAMLLTPTTAACNRDVERF	312
		1	2
Db	244	LPVTPVSTVAGSDSFLAALAHRLAOGGSVADAFRLGAAAGSKTLTVHGTOQLCGAEVERL	303
		1	2
Qy	313	F-ELAAP	319
		1	2
Db	304	YGVITIEP	311
		1	2

RESULT 6	063199 BURPS	PRELIMINARY;	PRT;	313 AA.
AC	063199 BURPS			
AD	063199 BURPS			
DT	25-OCT-2004 (TREMBlrel)	28, Created)		
DT	25-OCT-2004 (TREMBlrel)	28, Last sequence update)		
DT	25-OCT-2004 (TREMBlrel)	28, Last annotation update)		
DE	Purative 6-phosphotransferase.			
GN	OrderedLocName=BPSS1957;			
OS	Burkholderia pseudomallei (Pseudomonas pseudomallei).			
OC	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;			
OC	Burkholderiaceae; Burkholderia; pseudomallei group.			
OX	NCHI_TaxID=28450;			
RN	[1]			
RC	NUCLEOTIDE SEQUENCE.			
RP	STRAIN-K96243.			
RX	PubMed=15377794; DOI=10.1073/pnas.0403020101;			
RA	Holden M.T.G., Titchall R.W., Peacock S.U.C., Cerdano-Tarraga A.-M.,			
RA	Atkins T., Crossman I.C., Pitt T., Churcher C., Mungall K.L.,			
RA	Bentley S.D., Sebahia M., Thomson N.R., Baeson N., Beacham I.R.,			
RA	Brooke K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,			
RA	Chillingworth T., Cronin A., Crossset B., Davis P., Deshaizer D.,			
RA	Feilwell T., Fraser A., Hance Z., Hauser H., Holtroyd S., Jagsels K.,			
RA	Keith K.E., Madison M., Moule S., Price C., Quail M.A.,			
RA	Rabbinowitz E., Rutherford K., Sanders M., Simmonds M.,			
RA	Songsevitai S., Stevens K., Tunapa S., Vesaratchavee M.,			
RA	Whitehead S., Yeates C., Barrett B.G., Oyston P.C.F., Parthill J.;			
RT	"Genomic plasticity of the causative agent of melioidosis,			
RT	Burkholderia pseudomallei.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).			
RL	EMBL, BX571966; CAH39435.1; -, Genomic_DNA.			
DR	GO; GO:0016301; F:kinase activity; IEA.			
DR	GO; GO:0016740; F:transferase activity; IEA.			
DR	InterPro; IPR002173; PfKb.			
DR	InterPro; IPR011611; PfKb_region.			
DR	Pfam; PF00294; PfKb; 1.			
DR	PROSITE; PS00583; PFK_KINASES_1; 1.			
KW	Complete proteome; Kinase.			
SO	SEQUENCE 313 AA; 31941 MW; 2CE607B495039439 CRC64;			
Query Match	35.1%; Score 603.5; DB 2; Length 313;			
Best Local Similarity	42.2%; Pred. No. 3.9e-33;			
Matches 126; Conservative	59; Mismatches 115; Indels 1; Gaps 1			
QY	12 PRITTTTNPALDITTSVDVVRPTKMRGAFRYDPGGGGNNARIYVLTGGCTALPFA 71			
DB	2 PELVITLTLPNPAIDAVACVERVVDTRKRCGPARRDPGGGGNNARVLTTLGADSCAYYLA 61			
QY	72 GGSSTGLMALGADDAVPRPVPIAASSTESPTVNESPTAKOYRVLPGLPTVAEBOEC 131			
DB	62 GGGTGLALRGLADGVRAGHGDIDAGETRENSVLETSTGREFRVLDPGLAAHNEPRC 121			
QY	132 LDELRGAAASAAEVVAGSGLPGVADVYQYRVADI CRSSSTPLILDTSGGL-CHISSG 190			
DB	122 VVALGRLADSRVLTVNSGSLPGMPDDCYARLARPARSGVTVVDTGSPALAAALDAGV 181			
QY	191 FLKASVRELRECVGSELTLTEPQCLAAAHMLDRRAVVVVVSGSAGALLATGTHASHRF 250			
DB	182 YLVKPSLGEIRLALTGRLPLEDDGRLAAAVIAGVAGQIVALLTGLGDAALVSSDDAVRL 241			

QY 251 SSIPPTAAVGVAGAMVAALITVGLSRGSLIKSVLGNAAAGAAAMLITGTACNRDVE 310  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 242 PGVKAAVVASAIGADPSFVAGLVAAALNRRGANVADAARHMLAAPASASLSLTGTALGTREDLA 301  
  
QY 311 RFF 313  
| :  
Db 302 RLY 304

```

RESULT 7
062EC2_BURMA
ID 062EC2_BURMA PRELIMINARY; PRT; 313 AA.
AC 062EC2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, last annotation update)
DE 6-phosphofructokinase (EC 2.7.1.11).
GN Name=pfk; OrderedLocNames=BMAA0117;
OS Burkholderia mallei (Pseudomonas mallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxId=13373;
(1)
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 23344;
RX Pubmed=15377793; DOI=10.1073/pnas.0403306101;
RA Nieman W.C., Deshazer D., Kim H.S., Tetteelin H., Nelson K.E.,
RA Feldblyum T.C., Ulrich R.L., Konning C.M., Brintac L.M.,
RA Daugherty S.C., Davidson T.D., Deboy R.T., Dimitrov G., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Haft D.H., Khouri H.M., Kolonay J.F.,
RA Madupu R., Mohammed Y., Nelson W.C., Rauden P., Romero C.M.,
RA Sarría S., Selengut J., Shamblin C., Sullivan S.A., White O., Yu Y.,
RA Zafra N., Zhou L., Frazer C.M.;
RT "Structural flexibility in the Burkholderia mallei genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14246-14251(2004).
DR EMBL; CP000011; AAU6348.1; -, Genomic_DNA.
DR TIGR; BMAA0117; -.
DR GO; GO:0003872; F:6-phosphofructokinase activity; IEA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR002173; PFKB.
DR InterPro; IPR011611; PFKB_region.
DR Pfam; PF00294; PFKB_1.
DR PROSITE; PS00583; PFKB_KINASES_1; 1.
KW Complete proteome, kinase, transferase.
SQ SEQUENCE 313 AA; 31937 MW; 2C6607B49504F9F9 CRC64;

```

[illegible]

Db 302 R1Y 304

RESULT 8  
089189 BRAJA  
ID 089189 BRAJA PRELIMINARY; PRT; 295 AA.

AC 089189  
DT 01-JUN-2003 (Tremblrel. 24, Created)  
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE PFB family carbohydrate kinase.  
GN OrderedLocustNames=blx14659;  
OS Bradyrhizobium japonicum.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
NCBI\_TaxID=375;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=USDA 110;  
RX MEDLINE=2248498; PubMed=12597275;  
RA Kaneo T., Nakamura Y., Sato S., Minamisawa K., Uchiuni T.,  
RA Sasamoto S., Watanabe A., Idegawa K., Iiguchi M., Kawashima K.,  
RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,  
RA Tabata S.;  
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110.";  
RL DNA Res. 9:189-197(2002).  
DR EMBL; BA000040; BAC49924.1; -: Genomic DNA.  
DR GO; GO:0016301; F:kinase activity; IEA.  
DR InterPro; IPR002173; PFKB.  
DR InterPro; IPR011611; PFKB\_region.  
DR Pfam; PF00294; PFKB\_1.  
DR PROSITE; PS00583; PFKB\_KINASES\_1; 1.  
DR PROSITE; PS00584; PFKB\_KINASES\_2; UNKNOWN\_1.  
KM Complete proteome; Kinase.  
SQ SEQUENCE 295 AA; 30724 MW; E71C73B9308B3455 CRC64;

Query Match 33.4%; Score 574; DB 2; Length 295;  
Best Local Similarity 45.0%; Pred. No. 3.8e-31;  
Matches 127; Conservative 47; Mismatches 100; Indels 8; Gaps 3;

QY 34 PTEKRCGAPRYDGGGGINVARIYHVGCGSTALFPFGSGTSLMALLDAGVPPRYI 93  
DB 3 PYTKRCAPORDPGGGINVARIYHVGCGSTALFPFGSGTSLMALLDAGVPPRYI 62

QY 94 PIAASTRESFTVNSRTAKOYRFLVGPSTLVAEOCLDLRGAASAARVVASGSLRP 153  
DB 63 PISNDTRDITTFBTSKEQFLVPGACLNPEWQCLDAIARISPOAAVFIASGSLRP 122

QY 154 GVAADYQORVADICRRSST--PLIIDTSGGGLQ-HISSGVFLKASVREIRECVSEIL 209  
DB 123 GAPADYFGRVY---RSTGAAKVVDISGASIKRALEAGYVLIKPNRREFQELAGISGA 178

QY 210 TEPBQLAAHELDIGRAEVVVVSLGSGALLATRHASHRPSSIPMTAVSGVAGDAMVA 269  
DB 179 DESSLEGRRLIDRYRIEILALSMGPGALLTRFDILRANGPLLEPVSVSAGDSFLG 238

QY 270 AITVGLSGMSLISVRLGNAGAMLLTPGTAACNRDVER 311  
DB 239 AMVSRLANGDKLDSRLRYGVAGSAAALSPGTGLGADVHR 280

RESULT 9  
057PV8 SALCH  
ID 057PV8 SALCH PRELIMINARY; PRT; 310 AA.

AC 057PV8  
DT 10-MAY-2005 (Tremblrel. 30, Created)  
DT 10-MAY-2005 (Tremblrel. 30, Last sequence update)  
DT 10-MAY-2005 (Tremblrel. 30, Last annotation update)  
DE 6-phosphofructokinase II.  
GN Name=pfkfb; OrderedLocustNames=SC1347;  
OS Salmonella cholerae-suis (Salmonella enterica).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=591;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=SC-867;  
RX PubMed=15781495;  
RA Chiu C.-H., Tang P., Chu C., Hu S., Bao Q., Yu J., Chou Y.-Y.,  
RA Wang H.-S., Lee Y.-S.;  
RT "The genome sequence of Salmonella enterica serovar Choleraesuis, a  
highly invasive and resistant zoonotic pathogen.";  
RL Nucleic Acids Res. 33:1690-1698(2005).  
DR EMBL; AE017220; AAG5253.1; -: Genomic DNA.  
KM Complete proteome; Kinase.  
SQ SEQUENCE 310 AA; 32565 MW; 38C7EB972571F3F CRC64;

Query Match 31.4%; Score 539.5; DB 2; Length 310;  
Best Local Similarity 41.8%; Pred. No. 9e-29;  
Matches 127; Conservative 54; Mismatches 118; Indels 5; Gaps 4;

QY 13 RIITLTNMPALDITTSVDVVRPTEKRCGAPRYDGGGGINVARIYHVGCGSTALFPAG 72  
DB 3 RIVTLTAPSDSATITPQIYBEGLKCSAPVFEFGGGINVARAIAMLGATATAPAG 62

QY 73 GTSGLMALIGDGVFPVRIPIASTRESFTVNESTRAKOYRVLPGPSLTVAEQOCL 132  
DB 63 GATSEHVLALLADENVSTVDADKDWTRONLHVHVESSGQYRFVMEGATLDDEFRQ-L 121

QY 133 DELGAAASAFAVVASGSLPPGVADYVORVADICRRSSTPLIIDTSGGGLQHSIS--GV 190  
DB 122 EEOVLEIESGAILVSSSLPPGVVEKLTQLISAQOGKRCITDSSGDALITALAGDI 181

QY 191 FLKASVREIRECVSEILTEPEQL-AAHELDIGRAEVVVVSLGSGALLATRHASHR 249  
DB 182 ELVKNPKELISALVNRD-LTPDPDVRAQELVSGKARRVVSILGPGALIDISENCIO 240

QY 250 FSSIPMTAVSGVAGDAMVAITVGLSRGMSLISVRLGNAGAMLLTGTATACNRDV 309  
DB 241 VVPEPVVQSSTVGAGDSVMGAMTLKLAQDASLEBMYFVGAAGSAATLNOGTRLSHDT 300

QY 310 ERPF 313  
DB 301 QKIT 304

RESULT 10  
08ZPT5 SALTY  
ID 08ZPT5 SALTY PRELIMINARY; PRT; 310 AA.

AC 08ZPT5  
DT 01-MAR-2002 (Tremblrel. 20, Created)  
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE 6-phosphofructokinase II (EC 2.7.1.11).  
GN Name=pfkb; OrderedLocustNames=STM1326;  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
NCBI\_TaxID=602;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=LT2;  
RX MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan B., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
LT2.";  
RL EMBL; AE008757; AL20251.1; -: Genomic DNA.  
DR GO; GO:0003872; F:6-phosphofructokinase activity; IEA.  
DR GO; GO:0016301; F:kinase activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.





DR GO; GO:0016301; F:kinase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro: IPR002173; PfKB.  
 DR InterPro: IPR011611; PfKB\_region.  
 DR Pfam: PF00294; PfKB\_1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_1; 1.  
 DR PROSITE; PS00584; PFKB\_KINASES\_2; 1.  
 DR Complete proteome; Kinase.  
 KW SEQUENCE 310 AA; 32546 MW; 38CC7EB9724BED3F CRC64;  
 Query Match 31.1%; Score 534.5; DB 2; Length 310;  
 Best Local Similarity 41.4%; Pred. No. 2e-28;  
 Matches 126; Conservative 54; Mismatches 119; Indels 5; Gaps 4;  
 QY 13 RIITLTNNPALDITTSVDVVRPTKMKRCGARYDPGGGGINVARIVHVLGGCSALFPAG 72  
 DB 3 RIVTLTLPDLSATITTPQITPBGKRCAPVFEFGGGINVARIAHLGSGTAIRFPAG 62  
 QY 73 GSTSLMALIGDAGVPPFRVPIPAASTRESFTVNESRTAKQYRFLPGSLTVAEOGCL 132  
 DB 63 GATGEHLVALLADENVPSTYDADKMTKQNLHVHVESGEQYRFVMPGATLDDDEFRO-L 121  
 QY 133 DELRGAAASAAFPVVASGLPFGVAADYQYRVADICRSSTPLIDTSGGSLQHTSS--GV 190  
 DB 122 EGVLEIESGAILVLSGSLPGVKEVKLTQLISAQKQGRICIIDSSGDALTALALADI 181  
 QY 191 FLTKASVRELAECSGSELLTEPEOL--AAHLEIDRGAEVYVYVSGGALLARHSHR 249  
 DB 182 ELVFNLELSALVNRD-LTPQDDVRKAQGLVSGKRRRVVSLGSGALGIDSENQI 240  
 QY 250 FSSIPMTAVSGVAGDAAVAITVGLSRGSLIKSVRLGNAGAMLLTPTGAACNRDDV 309  
 DB 241 VVPPPVASQSVVGAAGDSVVGAMTLKLAQDASLEWVRGVAAGAATLNGTRICSHDT 300  
 QY 310 ERFP 313  
 DB 301 QKTY 304

RESULT 13  
 K6PF2\_ECOLI STANDARD; PRT; 309 AA.  
 AC P06999; P78065; P78260;  
 DT 01-APR-1998 (Rel. 07, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 13-SEP-2005 (Rel. 48, Last annotation update)  
 DE 6-phosphofructokinase isozyme 2 (EC 2.7.1.11) (Phosphofructokinase-2).  
 GN Name=pfkb; OrderedLocuNames=b1723;  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 NC NCI\_TaxID=562;  
 NX (1)  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=K12;  
 RC MEDLINE=84262485; PubMed=6235149; DOI=10.1016/0378-1119(84)90151-3;  
 RX Daidal F.;  
 RA "Nucleotide sequence of gene pfkb encoding the minor  
 RT phosphofructokinase of Escherichia coli K-12";  
 RL Gene 28:337-342(1984).  
 RN [2]  
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=K12 / MG1655.  
 RX MEDLINE=9742617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;  
 RA Blattner F.R.; Plunkett G. III; Bloch C.A.; Perna N.T.; Burland V.;  
 RA Riley M.; Collado-Vides J.; Glasner J.D.; Rode C.K.; Mayhew G.F.;  
 RA Gregor J.; Davis N.W.; Kirkpatrick H.A.; Goeden M.A.; Rose D.J.;  
 RA Mau B.; Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=K12;  
 RX MEDLINE=97251357; PubMed=9097039;  
 RA Alpha H., Babe T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,  
 RA Sempel G., Seki Y., Sivaundaram S., Tagami H., Takeda J.,  
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;  
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 28.0-40.1 min region on the linkage map";  
 RL DNA Res. 3:363-377(1996).  
 RN [4]  
 RN NUCLEOTIDE SEQUENCE OF 1-38.  
 RX MEDLINE=83294514; PubMed=6310120;  
 RA Daidal F.;  
 RT "Molecular cloning of the gene for phosphofructokinase-2 of  
 RT Escherichia coli and the nature of a mutation, pfkb1, causing a high  
 RT level of the enzyme";  
 RL J. Mol. Biol. 168:285-305(1983).  
 CC -1- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-fructose 1,6-bisphosphate.  
 CC -1- ENZYME REGULATION: PFK-2 is sensitive to inhibition by fructose 1,6-diphosphate.  
 CC -1- PATHWAY: Key control step of glycolysis.  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- MISCELLANEOUS: Only 10% of the activity present in the wild-type strain is phosphofructokinase-2.  
 CC -1- MISCELLANEOUS: This enzyme is not to be confused with 6-phosphofructo-2-kinase which is also called phosphofructokinase 2.  
 CC -1- MISCELLANEOUS: E.coli has two 6-phosphofructokinase enzymes.  
 CC -1- SIMILARITY: Belongs to the carbohydrate kinase pfkb family.  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.  
 CC EMBL; K02500; AAA24321.1; -; Genomic DNA.  
 DR EMBL; U00096; AAC74793.1; -; Genomic DNA.  
 DR EMBL; D90814; BAA15500.1; ALT\_INIT; Genomic DNA.  
 DR EMBL; D90815; BAA15506.1; ALT\_INIT; Genomic DNA.  
 DR EMBL; K00128; AAA24320.1; -; Genomic DNA.  
 DR PIR; C64931; KIECFB.  
 DR SWISS-2DPAGE; P06999; COLI.  
 DR EC02DBASE; E036.6; 6TH EDITION.  
 DR ECHOBASE; EB0694; -.  
 DR EcoGene; EG10700; pfkb.  
 DR InterPro; IPR002173; pfkb.  
 DR InterPro; IPR011611; PfKB\_region.  
 DR Pfam; PF00294; PfKB\_1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_1; 1.  
 DR PROSITE; PS00584; PFKB\_KINASES\_2; 1.  
 KW Complete proteome; Glycolysis; Kinase; Transferase.  
 FT FT  
 FT GKLCTAPVPEPG -> ENCAVHRCSNP (in Ref. 1 and 4).  
 FT  
 FT CONFLICT 155 171 AAKQGRICIVDSGGA -> LKKNGSAASSTVLGG (in Ref. 1).  
 FT  
 FT CONFLICT 245 246 PV -> AL (in Ref. 1).  
 FT  
 FT CONFLICT 257 258 SM -> RL (in Ref. 1).  
 FT  
 SQ SEQUENCE 309 AA; 32456 MW; A93BBE0D5801309 CRC64;  
 Query Match 30.9%; Score 531.5; DB 1; Length 309;  
 Best Local Similarity 41.1%; Pred. No. 3.2e-28;  
 Matches 123; Conservative 55; Mismatches 119; Indels 5; Gaps 4;  
 QY 13 RIITLTNNPALDITTSVDVVRPTKMKRCGARYDPGGGGINVARIVHVLGGCSALFPAG 72  
 DB 3 RIVTLTLPDLSATITTPQITPBGKRCAPVFEFGGGINVARIAHLGSGTAIRFPAG 62  
 QY 73 GSTSLMALIGDAGVPPFRVPIPAASTRESFTVNESRTAKQYRFLPGSLTVAEOGCL 132  
 DB 63 GATGEHLVALLADENVPSTYDADKMTKQNLHVHVESGEQYRFVMPGATLDDDEFRO-L 121







***This Page Blank (uspto)***



RESULT 2  
US-09-489-039A-10377  
; Sequence 10377, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 10377  
; LENGTH: 314  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-10377

Query Match 30.3%; Score 520.5; DB 2; Length 314;  
Best Local Similarity 39.6%; Pred. No. 8.2e-45;  
Matches 122; Conservative 52; Mismatches 131; Indels 3; Gaps 2;

QY 9 EGRRIITLTNNPALDITTSVDVVRPTEKMGCGAPRYDPGGGGINVARIYHVILGCGSTAL 68  
DB 3 EGMRTKTYTLTLPALDITTSVDVVRPTEKMGCGAPRYDPGGGGINVARIYHVILGCGSTAL 62  
QY 69 PPAAGSGTSLMALGDAAGVFRVPIPAASRSTFTVNESTAYQYFVLPGPSLTVAEQ 128  
DB 63 PPAAGSGTSLMALGDAAGVFRVPIPAASRSTFTVNESTAYQYFVLPGPSLTVAEQ 122  
QY 129 EOCDELRGAAASAFVVASGSLPPGVADYQYRVADICRRSSTPLIDTSGGLOH1-- 186  
DB 123 RR-LEEKVLTPEPSLVISSLPFGISVDLMQVKNAAQOGRCIIDSGDLAAALD 181  
QY 187 SSGVFLKASVRELRECVSELTTEPEQLAAAHLEIDRGAEEVVVVISGQALLATRHA 246  
DB 182 VGNIELVKPNQKELSAIVQRLSQPDVRLAAQSLISGKRVRVVSLPGGALGVDSG 241  
QY 247 SHRRSSIPMTAVSGVAGDAMVAATVGLSRGMSLISVRLGNAAAMLTPTAACNR 306  
DB 242 SVQVVPPEPKSQSTVAGDSWVGAMTURLAENASLEIDVRRGVAAASMAATTNGTRLCSR 301  
QY 307 DDVERFE 314  
DB 302 ANTOKTYD 309

RESULT 3  
US-09-134-000C-6381  
; Sequence 6381, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; FILE REFERENCE: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6381  
; LENGTH: 318  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-6381

Query Match 18.7%; Score 322; DB 2; Length 318;  
Best Local Similarity 29.6%; Pred. No. 1.9e-24;

Matches 85; Conservative 51; Mismatches 141; Indels 10; Gaps 4;  
QY 14 ITLTNNPALDITTSVDVVRPTEKMGCGAPRYDPGGGGINVARIYHVILGCGSTALP 73  
DB 7 IVTVTPNSIDISVLDLHKLDTVNRFTSQYTKTPGGKGLVTVFVHDLGADVATGVLGG 66  
QY 74 STGSLMALGDAAGVFRVPIPAASRSTFTVNESTAYQYFVLPGPSLTVAEQECLD 133  
DB 67 FHGAFINELKKNNIPAFSTISKEERTDSIAL--LHEGNTTEILEAGPTVSPESISNPLE 124  
QY 134 ELRGAAASAFVVASGSLPPGVADYQYRVADICRRSSTPLIDTSGGLOHISG---V 190  
DB 125 NPDQLKQAEIVITISGLAGLSPDFYQELVQKAAHAEVVLDTGSDSLRQLQGPWK 184  
QY 191 FLKASVRELRECVSELTTEPEQLAAAHLEIDR--GRAEEVVVVISGQALLATRHA 247  
DB 185 YLKPNELEELGQDFSEN--LAAVQYALTKPMPAGIEWIVISIGKGAIALKHNDQ 242  
QY 248 HRSSSIPTMTAVSGVAGDAMVAATVGLSRGMSLISVRLGNAAAMLTPTAACNR 294  
DB 243 YRVKIPITQAKNPVSGDATTIAGLADAPABAILKWMAGMA 289

RESULT 4  
US-09-134-000C-3807  
; Sequence 3807, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; FILE REFERENCE: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3807  
; LENGTH: 321  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-3807

Query Match 17.3%; Score 298; DB 2; Length 321;  
Best Local Similarity 29.5%; Pred. No. 5.6e-22;  
Matches 88; Conservative 59; Mismatches 139; Indels 12; Gaps 6;

QY 10 GKPRITLTNNPALDITTSVDVVRPTEKMGCGAPRYDPGGGGINVARIYHVILGCGSTALP 69  
DB 14 GKQMTYTVTLNPSIDISVLDLHKLDTVNRFTSQYTKTPGGKGLVTVFVHDLGADVATGVLGG 73  
QY 70 PPAAGSGTSLMALGDAAGVFRVPIPAASRSTFTVNESTAYQYFVLPGPSLTVAEQE 129  
DB 74 FLGGFTGSPFTADWLKKEELQTNFTPSADRLNIKU--KSDTELEINGL--GPAALTVEE-- 128  
QY 130 OCDELRGAAASAFVVASGSLPPGVADYQYRVADICRRSSTPLIDTSGGLOH1-- 185  
DB 129 --IOLKQAVSRQAGDIVLSSSTASLAKGYEELIQVKEKGAEPVIDTIGEDLMNA 186  
QY 186 ISSGVFLKASVRELRECVSELTTEPEQLAAAHLEIDRGAEEVVVVISGQALLATRHA 245  
DB 187 LSQKPLLVKNNHLEALVHTFTSIEDLIPYGRILLEEG--AQHVILSMAGDALFTTE 245  
QY 246 ASHRFSIPTMTAVSGVAGDAMVAATVGLSRGMSLISVRLGNAAAMLTPTAACNR 303  
DB 246 GYVRSNVLKRLPKNSVAGDSMTAGFTGNFSKTQDPLEAFKMWAGCAYATAFSDDLAS 303

RESULT 5  
US-09-134-001C-5583  
; Sequence 5583, Application US/09134001C  
; Patent No. 6380370

```

; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5583
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5583

Query Match      16.5%; Score 284; DB 2; Length 312;
Best Local Similarity 26.8%; Pred. No. 1.5e-20;
Matches 80; Conservative 68; Mismatches 141; Indels 10; Gaps 6;

QY 14 IITLNNPALDITTSVDVVRPTEKMGCGAPRYDPPGGGINVARIYHVLGGCTALFPAGG 73
DB 8 IYTVNPSIDYIITDGFELISGLNRAKATYKFAAGKGINVSKVLKTLVDVSTALGSGG 67
QY 74 STGSLMALGADAGVPRVPIAASTRESFTVN--ESRTAKQYRFVLPGPSLTVAEOECL 132
DB 68 PFGGFIAGTLEDSNIQSDPFGVDEDT---INVKLKSGSGTEINAGPKYTHAFOEDL 123
QY 133 DELRGAAASAAFFVAGSGLPPGVADYQYRVADICRBSSTPLIIDTSGGGLQHI--SSGVF 191
DB 124 SQIR-RTTNDIVIVAGSVSPNSIPSDAYAQIAQITKGTGLVDADKDLVETVLPPYRPL 182
QY 192 LKASVRELRECVSELTEPEOLAANHELIDRGARAVVVSOGALLATRHASHRES 251
DB 183 FIKKXDELEMYENTIVKASDEVDIKYKEIKLKG--AGSVIISLGGDGLIYDQHSIKAV 241
QY 252 SIPTAVSGVAGDAMVAATVGLSRGMSLIKSYRLNAGAAAMLLTPGTAAACNRDDVE 310
DB 242 NPGSHVNVTVSGSDSTVAGVAGVAGSLMGLNIDEAFOAVASGTATAFSEDLA--TRDAIE 298

RESULT 6
US-09-715-040-2
; Sequence 2, Application US/09715040
; Patent No. 6806068
; GENERAL INFORMATION:
; APPLICANT: Mockel, Bettina
; APPLICANT: Pfeifferle, Walter
; TITLE OF INVENTION: NEW NUCLEOTIDE SEQUENCES WHICH ENCODE THE PFK GENE
; FILE REFERENCE: 21123/275321
; CURRENT APPLICATION NUMBER: US/09/715,040
; CURRENT FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-715-040-2

Query Match      16.2%; Score 279; DB 2; Length 330;
Best Local Similarity 31.2%; Pred. No. 5.2e-20;
Matches 102; Conservative 46; Mismatches 133; Indels 46; Gaps 12;

QY 14 IITLNNPALDITTSVDVVRPTEKMGCGAPRYD-----PGGGGINVARIYHVLGGCTAL 68
DB 2 IITFPNPSIDSTSLG-----EELSRGVSQRLDSVTAVAGGKGINVAHAVALAFETLAV 57
QY 69 FPAAGSTGSLMALGADAGVPRVPIAASTRESFTVNE--SRTAKQYRFVLPGPSLTVA 126
DB 58 FPAAGL--DPFVPLVRDIGLPEVETVINKVNTITVTEPDGTTK-----LNGPAPLS 110

; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
```

```

QY 127 EQ-----EQLDELRGAAASAAFFVAGSGLPPGVADYQYRVADICR--RSSTPLIIDT 178
DB 111 EQLKRLSEKVLIDALR---PEVTWVVLVAGSLPGCAPVDVYARLTALIHSAKRPVRAVD 167
QY 179 S-----GGGLQHSISGVLLKASVRELRECV---GSELLTEPEQ-----LAAAHLEI 222
DB 168 SDRPLMALGESLTPGAPVPLIKRNGLELGQLANTDGELEBARAAGDYDAI1AAADV 227
QY 223 DRGRAEVVVSLSOGALLATRHASHRESFISIPMTAVSGVAGDAMVAATVGLSRGMSLI 282
DB 228 NRG-IEQVLTVLGAAGAVLVNAGCAWTATSPKIDVSTVAGDCALAGFMAASQKKTLE 286
QY 283 KSVRLNAGAAAMLLTPGTAAACNRDDV 309
DB 287 ESLLNAVSYGSTAASLPGTITPRPDQL 313

RESULT 7
US-10-098-626-4
; Sequence 4, Application US/10098626
; Patent No. 6921651
; GENERAL INFORMATION:
; APPLICANT: Patrick, Mike, et al.
; TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY USING CORYNEFORM I
; FILE REFERENCE: 032301 WN 269
; CURRENT APPLICATION NUMBER: US/10/098,626
; CURRENT FILING DATE: 2002-03-18
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-10-098-626-4

Query Match      16.2%; Score 279; DB 2; Length 330;
Best Local Similarity 31.2%; Pred. No. 5.2e-20;
Matches 102; Conservative 46; Mismatches 133; Indels 46; Gaps 12;

QY 14 IITLNNPALDITTSVDVVRPTEKMGCGAPRYD-----PGGGGINVARIYHVLGGCTAL 68
DB 2 IITFPNPSIDSTSLG-----EELSRGVSQRLDSVTAVAGGKGINVAHAVALAFETLAV 57
QY 69 FPAAGSTGSLMALGADAGVPRVPIAASTRESFTVNE--SRTAKQYRFVLPGPSLTVA 126
DB 58 FPAAGL--DPFVPLVRDIGLPEVETVINKVNTITVTEPDGTTK-----LNGPAPLS 110
QY 127 EQ-----EQLDELRGAAASAAFFVAGSGLPPGVADYQYRVADICR--RSSTPLIIDT 178
DB 111 EQLKRLSEKVLIDALR---PEVTWVVLVAGSLPGCAPVDVYARLTALIHSAKRPVRAVD 167
QY 179 S-----GGGLQHSISGVLLKASVRELRECV---GSELLTEPEQ-----LAAAHLEI 222
DB 168 SDRPLMALGESLTPGAPVPLIKRNGLELGQLANTDGELEBARAAGDYDAI1AAADV 227
QY 223 DRGRAEVVVSLSOGALLATRHASHRESFISIPMTAVSGVAGDAMVAATVGLSRGMSLI 282
DB 228 NRG-IEQVLTVLGAAGAVLVNAGCAWTATSPKIDVSTVAGDCALAGFMAASQKKTLE 286
QY 283 KSVRLNAGAAAMLLTPGTAAACNRDDV 309
DB 287 ESLLNAVSYGSTAASLPGTITPRPDQL 313

; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
```



Db 64 GGALGQYIEBOIEETTRIKQAFKIKGETRNCIAI--LHEGOQTEILEKGPTEILEKESEEF 121  
Qy 132 LDELKGAASAAAFVVASGSLPVGVAADYYGRVADICRRSSTPLIDTSGGGLQHI-SGGV 190  
Db 122 KSHLLKPKENDVAVMSSGLPKGINTDYADIVRLAKEGGLITLIDSSGSGLEEVLSNV 181  
Qy 191 --PLIKASVRELRECVSGSELTEPEQLAAHELIDRGRAEVVVSLSGQALLATRHASH 248  
Db 182 KPTVICKNIDELSQLINVKVINDIKELKAAVSOPIFNDIEWIIVSLGSEGAFAKHQKFX 241  
Qy 249 RFSISPTMAVSGVAGDAMVAITVGLSRGWSLKSVALGNAAGAMLTGTGAACRRD 308  
Db 242 KVNIPNIKVVAVSGSDTVAGIASGLIHQOTDELLKKAAFGMLNAMEOQTHINTDK 301  
Qy 309 VERFFE 314  
Db 302 FDEIFK 307  
RESULT 11  
US-09-107-433-4601  
; Sequence 4601, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNO  
; THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: <Unknown>  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: <Unknown>  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107.433  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/ 085131  
; FILING DATE: May 12, 1998  
; APPLICATION NUMBER: 60/051553  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arinello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 4601:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 310 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULAR TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...310  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4601:  
US-09-107-433-4601  
Query Match 15.1%, Score 260, DB 2, Length 310,  
Best Local Similarity 28.0%, Pred. No. 4.2e-18,

Matches 81; Conservative 51; Mismatches 137; Indels 20; Gaps 7;  
Qy 12 PRITLTNMPALDITTSVDVVRPTKRRCGAPRYDPGGGGINVARIYHVLGGCSTALFPA 71  
Db 1 PMILTVMNNSIDISYLDLDELKIDTVNRVVDVTKTAGKGLNTRVLSEEGDSVLATGLV 60  
Qy 72 GCGTSGLSLMLLGDAGYFFRVVPIASTRESFTV---NESRTRAKYRFVLPGSLTVAE 127  
Db 61 GSKGEFLVHIHQD--VKDQFPSIOGETRNCIILHNDQTEVLEK-----GEVLEOE 113  
Qy 128 QEOGLDELKGAASAAAFVVASGSLPVGVAADYYGRVADICRRSSTPLIDTSGGGLQHI 187  
Db 114 GQDFLEHFKQLLSVEVAISGLPAGLPDYTASLVELANQAKLVVLDSCGALQAVL 173  
Qy 188 SGVF---LTKASVRELRECVSGSELTEPEQL-AAHELIDRGRAEVVVSLSGQALLAT 243  
Db 174 ESPHKPTVINKNNEELSGLLGRVSEDELLEKVELGSLFTG-IEWIIVSLGANGTAKH 232  
Qy 244 RHASHRFSISPTMAVSGVAGDAMVAITVGL-----SRGWSLKSVALG 288  
Db 233 GDTFFYKVDIPRIQVNVFVSGSDTVAGISSGLHKESDAGLLIKANVLG 281  
RESULT 12  
US-08-961-539-2  
; Sequence 2, Application US/08961539  
; Patent No. 5861281  
; GENERAL INFORMATION:  
; APPLICANT: Zalacaln, Magdalena  
; APPLICANT: Brown, James R.  
; TITLE OF INVENTION: No. 5861281el 1acc  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert, Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2793  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961.539  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dickinson, Q. Todd  
; REGISTRATION NUMBER: 28,354  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215/994-2252  
; TELEFAX: 215/994-2222  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 303 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-961-539-2  
Query Match 14.0%, Score 240, DB 1, Length 303,  
Best Local Similarity 27.4%, Pred. No. 4.5e-16;  
Matches 78; Conservative 54; Mismatches 141; Indels 12; Gaps 5;  
Qy 14 ITLTNMPALDITTSVDVVRPTKRRCGAPRYDPGGGGINVARIYHVLGGCSTALFPA 73



```
Db 2 IYVTLNPSIDYIVRLQOVKGVNVRMDSDDKFKAGGKINVRVLKRLNISNTATGFIGG 61
Qy 74 STGSLMALIGDAGVPRVPIAASSTRESFTVNESRTAKOYRVLPGPSLTVAEQOCLD 133
Db 62 FTGKFITDITLAEIEIETRFQVADTRINVKIKADQETE-----INGTGPVPEVK--LE 114
Qy 134 ELRGAAASAA--FVVASGSLPPGVADYQVRVADICRSSTPLILDTSGGGL-QHISSG 189
Db 115 ELKAILSSLTAEDTVVFASSAKNLGNVIYKDLISLRTQTGAQVCDPFEQOTLIDSLDYQ 174
Qy 190 VFLKASVRELRECVSESLITEPEQLAAHEDLRGAEVVYVSLSGQALLATRRASHR 249
Db 175 PLLVKPNHNLGALFGVKLESDEIEKYARELLAKG-AQNVITISMAGDGLVTSSEGAYF 233
Qy 250 FSSIPTAVSGVAGDAMVAITVGLSRGWSLIKSVRLGNAGAA 294
Db 234 AKPIKGTIVKNSVGAQDSMVAGFTGEFVKSDVAEAFKMGVACGTA 278

RESULT 13
US-09-185-826-2
; Sequence 2, Application US/09185826
; Patent No. 6171840
; GENERAL INFORMATION:
; APPLICANT: Zalecain, Magdalena
; TITLE OF INVENTION: No. 6171840e1 1acc
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/185.826
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/961.539
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Q. Todd
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10114
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215/994-2252
; TELEFAX: 215/994-2222
; TEXT:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-185-826-2

Query Match 14.0%; Score 240; DB 2; Length 303;
Best local Similarity 27.4%; Pred. No. 4.5e-16;
Matches 78; Conservative 54; Mismatches 141; Indels 12; Gaps 5;

Qy 14 IITLTNNPALDITTSVDVVRPTEKRCGAPRYDPGGGGINVARIHVHVGSCSTALFPAGG 73
Db 2 IYVTLNPSIDYIVRLQOVKGVNVRMDSDDKFKAGGKINVRVLKRLNISNTATGFIGG 61
Qy 74 STGSLMALIGDAGVPRVPIAASSTRESFTVNESRTAKOYRVLPGPSLTVAEQOCLD 133
Db 62 FTGKFITDITLAEIEIETRFQVADTRINVKIKADQETE-----INGTGPVPEVK--LE 114
Qy 134 ELRGAAASAA--FVVASGSLPPGVADYQVRVADICRSSTPLILDTSGGGL-QHISSG 189
Db 115 ELKAILSSLTAEDTVVFASSAKNLGNVIYKDLISLRTQTGAQVCDPFEQOTLIDSLDYQ 174
Qy 190 VFLKASVRELRECVSESLITEPEQLAAHEDLRGAEVVYVSLSGQALLATRRASHR 249
Db 175 PLLVKPNHNLGALFGVKLESDEIEKYARELLAKG-AQNVITISMAGDGLVTSSEGAYF 233
Qy 250 FSSIPTAVSGVAGDAMVAITVGLSRGWSLIKSVRLGNAGAA 294
Db 234 AKPIKGTIVKNSVGAQDSMVAGFTGEFVKSDVAEAFKMGVACGTA 278

RESULT 15
US-09-107-433-4423
; Sequence 4423, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
```

```
Db 62 FTGKFITDITLAEIEIETRFQVADTRINVKIKADQETE-----INGTGPVPEVK--LE 114
Qy 134 ELRGAAASAA--FVVASGSLPPGVADYQVRVADICRSSTPLILDTSGGGL-QHISSG 189
Db 115 ELKAILSSLTAEDTVVFASSAKNLGNVIYKDLISLRTQTGAQVCDPFEQOTLIDSLDYQ 174
Qy 190 VFLKASVRELRECVSESLITEPEQLAAHEDLRGAEVVYVSLSGQALLATRRASHR 249
Db 175 PLLVKPNHNLGALFGVKLESDEIEKYARELLAKG-AQNVITISMAGDGLVTSSEGAYF 233
Qy 250 FSSIPTAVSGVAGDAMVAITVGLSRGWSLIKSVRLGNAGAA 294
Db 234 AKPIKGTIVKNSVGAQDSMVAGFTGEFVKSDVAEAFKMGVACGTA 278

RESULT 14
US-09-583-110-5072
; Sequence 5072, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PAT000-07A
; CURRENT APPLICATION NUMBER: US/09/583.110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107.433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085.131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051.553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 5072
; LENGTH: 303
; TYPE: PRN
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (17)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-09-583-110-5072

Query Match 13.4%; Score 230; DB 2; Length 303;
Best local Similarity 27.0%; Pred. No. 4.8e-15;
Matches 77; Conservative 53; Mismatches 143; Indels 12; Gaps 5;

Qy 14 IITLTNNPALDITTSVDVVRPTEKRCGAPRYDPGGGGINVARIHVHVGSCSTALFPAGG 73
Db 2 IYVTLNPSIDYIVRLQOVKGVNVRMDSDDKFKAGGKINVRVLKRLNISNTATGFIGG 61
Qy 74 STGSLMALIGDAGVPRVPIAASSTRESFTVNESRTAKOYRVLPGPSLTVAEQOCLD 133
Db 62 FTGKFITDITLAEIEIETRFQVADTRINVKIKADQETE-----INGTGPVPEVK--LE 114
Qy 134 ELRGAAASAA--FVVASGSLPPGVADYQVRVADICRSSTPLILDTSGGGL-QHISSG 189
Db 115 ELKAILSSLTAEDTVVFASSAKNLGNVIYKDLISLRTQTGAQVCDPFEQOTLIDSLDYQ 174
Qy 190 VFLKASVRELRECVSESLITEPEQLAAHEDLRGAEVVYVSLSGQALLATRRASHR 249
Db 175 PLLVKPNHNLGALFGVKLESDEIEKYARELLAKG-AQNVITISMAGDGLVTSSEGAYF 233
Qy 250 FSSIPTAVSGVAGDAMVAITVGLSRGWSLIKSVRLGNAGAA 294
Db 234 AKPIKGTIVKNSVGAQDSMVAGFTGEFVKSDVAEAFKMGVACGTA 278

RESULT 15
US-09-107-433-4423
; Sequence 4423, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
```



***This Page Blank (uspto)***

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 06:14:33 ; Search time 71.2896 Seconds  
(without alignments)  
1986.885 Million cell updates/sec

Title: US-10-617-038-22

Perfect score: 1720  
Sequence: 1 MTEPAAWDEKPRITLITM.....TEVGQDYVHPIVNPASP 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1720	100.0	339	US-10-282-122A-62568	Sequence 62568, A
2	1720	100.0	339	US-10-282-122A-64633	Sequence 64633, A
3	1720	100.0	339	US-10-617-038-22	Sequence 22, App1
4	726	42.2	310	US-10-369-493-19155	Sequence 19155, A
5	617.5	35.9	308	US-10-369-493-10209	Sequence 10209, A
6	587.5	34.7	505	US-10-282-122A-50578	Sequence 50578, A
7	587.5	34.2	338	US-10-282-122A-48067	Sequence 48067, A
8	534.5	31.1	310	US-09-815-242-13976	Sequence 13976, A
9	531.5	30.9	309	US-09-815-242-10180	Sequence 10180, A
10	531.5	30.9	309	US-10-369-493-851	Sequence 851, App
11	531.5	30.9	309	US-10-282-122A-55670	Sequence 55670, A
12	531.5	30.9	311	US-10-282-122A-55963	Sequence 55963, A
13	355	20.6	317	US-10-156-761-11226	Sequence 11226, A
14	330	19.2	309	US-10-156-761-11476	Sequence 11476, A
15	311	18.1	303	US-10-282-122A-45630	Sequence 45630, A
16	304	17.7	309	US-10-282-122A-52089	Sequence 52089, A
17	302	17.6	1249	US-10-450-763-52025	Sequence 52025, A
18	294	17.1	304	US-10-501-282-262	Sequence 262, App
19	290	16.9	305	US-09-815-242-10787	Sequence 10787, A
20	289	16.8	304	US-10-873-467-74	Sequence 74, App1
21	289	16.8	305	US-10-282-122A-57069	Sequence 57069, A
22	284	16.5	306	US-10-282-122A-70667	Sequence 70667, A
23	284	16.5	312	US-10-724-972A-7339	Sequence 7339, App
24	283	16.5	306	US-10-282-122A-44238	Sequence 44238, A
25	282	16.4	303	US-09-815-242-5393	Sequence 5393, App
26	282	16.4	306	US-09-815-242-12300	Sequence 12300, A
27	279.5	16.2	312	US-10-501-282-1494	Sequence 1494, App

28	279	16.2	330	US-09-738-626-5624	Sequence 5624, App
29	279	16.2	330	US-10-098-626-4	Sequence 4, App1
30	279	16.2	330	US-10-781-014-58	Sequence 58, App1
31	279	16.2	330	US-10-494-836-58	Sequence 58, App1
32	276.5	16.1	302	US-09-815-242-4955	Sequence 4955, App
33	276.5	16.1	308	US-09-815-242-10775	Sequence 10775, A
34	276.5	16.1	308	US-10-282-122A-42555	Sequence 42555, A
35	272	15.8	304	US-10-282-122A-57485	Sequence 57485, A
36	270	15.7	303	US-10-282-122A-45645	Sequence 45645, A
37	269.5	15.7	304	US-10-282-122A-59615	Sequence 59615, A
38	266	15.5	306	US-10-369-493-17156	Sequence 17156, A
39	266	15.5	307	US-10-282-122A-60536	Sequence 60536, A
40	264.5	15.4	314	US-10-724-972A-6354	Sequence 6354, App
41	264	15.3	306	US-10-282-122A-71925	Sequence 71925, A
42	263.5	15.3	307	US-10-282-122A-47216	Sequence 47216, A
43	260	15.1	310	US-10-617-320-4601	Sequence 4601, App
44	252.5	14.7	321	US-10-617-320-4601	Sequence 53758, A
45	250	14.5	309	US-10-472-928-2382	Sequence 2382, App

#### ALIGNMENTS

RESULT 1  
US-10-282-122A-62568  
Sequence 62568, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Olesen, Karl  
APPLICANT: Zykkind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: EUTPA-034A  
CURRENT FILING DATE: 2003-02-20  
PRIOR FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 62568  
LENGTH: 339  
TYPE: PRT  
ORGANISM: Mycobacterium bovis  
US-10-282-122A-62568

Query Match 100.0%; Score 1720; DB 4; Length 339;  
Best Local Similarity 100.0%; Pred. No. 7.1e-154;  
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEPAAWDEGKPRITITLTMNPAIDITTSVDVVRPTEKMRGAPRYDPGGGINVARIYHV 60  
DB 1 MTEPAAWDEGKPRITITLTMNPAIDITTSVDVVRPTEKMRGAPRYDPGGGINVARIYHV 60

QY 61 LGGCSTALFPAGSGTSLMALLDGAGVPRVPIAATSTRESFTVNESRTAKQYRFLVPG 120  
DB 61 LGGCSTALFPAGSGTSLMALLDGAGVPRVPIAATSTRESFTVNESRTAKQYRFLVPG 120

QY 121 PSLTVAEOEQCLDELKGAASAAAFVVASGSLPGVADYQYVADICRSTPLILDTSG 180  
DB 121 PSLTVAEOEQCLDELKGAASAAAFVVASGSLPGVADYQYVADICRSTPLILDTSG 180

QY 181 GGLQHTSSGVFLKASVRELRCEVSELTPEQLAAHEILDRGRAEVVVSIGSGAL 240  
DB 181 GGLQHTSSGVFLKASVRELRCEVSELTPEQLAAHEILDRGRAEVVVSIGSGAL 240

QY 241 LATRHASHRFSIIPMTAVSGVAGDAMVAALITVGLSRGMSLIKSVRLGNAGAAMLTPG 300  
DB 241 LATRHASHRFSIIPMTAVSGVAGDAMVAALITVGLSRGMSLIKSVRLGNAGAAMLTPG 300

QY 301 TAAACNRDVERFFELAAPEFTEVGDDQYVWHPVYNPEASP 339  
DB 301 TAAACNRDVERFFELAAPEFTEVGDDQYVWHPVYNPEASP 339

## RESULT 2

US-10-282-122A-64633  
Sequence 64633, Application US/10282122A  
Publication No. US20040029129A1

## GENERAL INFORMATION:

APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zykend, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1

SEQ ID NO 64633  
LENGTH: 339  
TYPE: PRT  
ORGANISM: Mycobacterium tuberculosis  
US-10-282-122A-64633

Query Match 100.0%; Score 1720; DB 4; Length 339;  
Best Local Similarity 100.0%; Pred. No. 7.1e-154;  
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEPAAWDEGKPRITITLTMNPAIDITTSVDVVRPTEKMRGAPRYDPGGGINVARIYHV 60  
DB 1 MTEPAAWDEGKPRITITLTMNPAIDITTSVDVVRPTEKMRGAPRYDPGGGINVARIYHV 60

QY 61 LGGCSTALFPAGSGTSLMALLDGAGVPRVPIAATSTRESFTVNESRTAKQYRFLVPG 120  
DB 61 LGGCSTALFPAGSGTSLMALLDGAGVPRVPIAATSTRESFTVNESRTAKQYRFLVPG 120

QY 121 PSLTVAEOEQCLDELKGAASAAAFVVASGSLPGVADYQYVADICRSTPLILDTSG 180  
DB 121 PSLTVAEOEQCLDELKGAASAAAFVVASGSLPGVADYQYVADICRSTPLILDTSG 180

QY 181 GGLQHTSSGVFLKASVRELRCEVSELTPEQLAAHEILDRGRAEVVVSIGSGAL 240  
DB 181 GGLQHTSSGVFLKASVRELRCEVSELTPEQLAAHEILDRGRAEVVVSIGSGAL 240

QY 241 LATRHASHRFSIIPMTAVSGVAGDAMVAALITVGLSRGMSLIKSVRLGNAGAAMLTPG 300  
DB 241 LATRHASHRFSIIPMTAVSGVAGDAMVAALITVGLSRGMSLIKSVRLGNAGAAMLTPG 300

QY 301 TAAACNRDVERFFELAAPEFTEVGDDQYVWHPVYNPEASP 339  
DB 301 TAAACNRDVERFFELAAPEFTEVGDDQYVWHPVYNPEASP 339

## RESULT 3

US-10-617-038-22  
Sequence 22, Application US/10617038  
Publication No. US20040057963A1

## GENERAL INFORMATION:

APPLICANT: Andersen, Peter  
APPLICANT: Rosenkrands, Ida  
APPLICANT: Scrym, Anette  
TITLE OF INVENTION: Therapeutic TB Vaccine  
FILE REFERENCE: SS15AUSA  
CURRENT APPLICATION NUMBER: US/10/617,038  
CURRENT FILING DATE: 2003-07-11  
PRIOR APPLICATION NUMBER: DK PA 2002 01098  
PRIOR FILING DATE: 2002-07-13  
PRIOR APPLICATION NUMBER: US 60/401,725  
PRIOR FILING DATE: 2002-08-07  
NUMBER OF SEQ ID NOS: 187  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 22  
LENGTH: 339  
TYPE: PRT  
ORGANISM: Mycobacterium tuberculosis  
US-10-617-038-22

Query Match 100.0%; Score 1720; DB 4; Length 339;  
Best Local Similarity 100.0%; Pred. No. 7.1e-154;  
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEPAAWDEGKPRITITLTMNPAIDITTSVDVVRPTEKMRGAPRYDPGGGINVARIYHV 60  
DB 1 MTEPAAWDEGKPRITITLTMNPAIDITTSVDVVRPTEKMRGAPRYDPGGGINVARIYHV 60

QY 61 LGGCSTALFPAGSGTSLMALLDGAGVPRVPIAATSTRESFTVNESRTAKQYRFLVPG 120  
DB 61 LGGCSTALFPAGSGTSLMALLDGAGVPRVPIAATSTRESFTVNESRTAKQYRFLVPG 120

QY 121 PSLTVAEOEQCLDELKGAASAAAFVVASGSLPGVADYQYVADICRSTPLILDTSG 180  
DB 121 PSLTVAEOEQCLDELKGAASAAAFVVASGSLPGVADYQYVADICRSTPLILDTSG 180

Db 121 PSLTVAEOEQCLDELRGAAAFVVASGSLPPGVADYORVADICRRSSTPLILDTSG 180  
Qy 181 GGLHISSGVFLTKASVRELRECVSELTTEPEQLAAHELIDRGAEVVVSLGSGAL 240  
Db 181 GGLHISSGVFLTKASVRELRECVSELTTEPEQLAAHELIDRGAEVVVSLGSGAL 240  
Qy 241 LATRHASHRFSISIPMTAVSGVAGADAMVAATVGLSRGWSLIKSVRLGNAAGAMLTPG 300  
Db 241 LATRHASHRFSISIPMTAVSGVAGADAMVAATVGLSRGWSLIKSVRLGNAAGAMLTPG 300  
Qy 301 TAACNRDVERREFELAAPETEVGDOQYWMHPVNPPEAP 339  
Db 301 TAACNRDVERREFELAAPETEVGDOQYWMHPVNPPEAP 339

## RESULT 4

US-10-369-493-19155  
; Sequence 19155, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 19155  
; LENGTH: 310  
; TYPE: PRT  
; ORGANISM: Myxococcus xanthus  
US-10-369-493-19155

Query Match 42.2%; Score 726; DB 4; Length 310;  
Best Local Similarity 49.0%; Pred. No. 7,1e-60;  
Matches 152; Conservative 56; Mismatches 100; Indels 2; Gaps 2;

Qy 12 PRITLTNNPALDITTSVDVVRPTKRCGAPRYDPGGGGINVARIYHVLGGCTALPFA 71  
Db 1 PHITLTNNPALDITTSVDVVRPTKRCGAPRYDPGGGGINVARIYHVLGGCTALPFA 71  
Qy 72 GGSSTGLMALLDGAGVPRVPIAASSTRESFTVNESRTAKQYRFVLPGPSLTVAEOEQ 131  
Db 61 GGAGSGLLELLEKGLRRSVPLODNRRESFTVAEGTGAEBYRFLPGPTLSREMQRC 120  
Qy 132 LDELRGAAAFVVASGSLPPGVADYORVADICRRSSTPLILDTSGGGLQ-HISSGV 190  
Db 121 LDALIEDVAVGATYIVASGSLAPGVPEDFYARVARGLKLAIRVVVDITGPPRLAALEEGV 180  
Qy 191 FLTKASVRELRECVSELTTEPEQLAAHELIDRGAEVVVSLGSGALLATRHASHRF 250  
Db 181 FMAAPNRRELKDLGASVEDPATQAAARBELVAQGRLEVLAVSHGADALVTTTQAQFRA 240  
Qy 251 SSIPMTAVSGVAGADAMVAATVGLSRGWSLIKSVRLGNAAGAMLTPGTAACNRDDVE 310  
Db 241 TPPVRITVSSYGAGDSFVAGMTLALARGOSTEDALRWGIAAGTALTLGQIDLCYRADVE 300  
Qy 311 RFF-ELAAEP 319  
Db 301 RLFTQVKAEP 310

RESULT 5  
US-10-369-493-10209  
; Sequence 10209, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 10209  
; LENGTH: 308  
; TYPE: PRT  
; ORGANISM: Cytophaga hutchinsonii  
US-10-369-493-10209

Query Match 35.9%; Score 617.5; DB 4; Length 308;  
Best Local Similarity 43.0%; Pred. No. 1,3e-49;  
Matches 131; Conservative 55; Mismatches 118; Indels 1; Gaps 1;

Qy 11 KPRIITLTNNPALDITTSVDVVRPTKRCGAPRYDPGGGGINVARIYHVLGGCTALPFA 70  
Db 2 KSIIFLTLSLPADFCKSTSVESVQPEHKLRCDSPIYCGGGGINVSRALKKLNGSSVALYP 61  
Qy 71 AGSSTGLMALLDGAGVPRVPIAASSTRESFTVNESRTAKQYRFVLPGPSLTVAEOEQ 130  
Db 62 IGGDTGNHFRALLTEBIEQITFOIKWTKRENTVRSSTNNDFRFGMPSESEYKKEWEQ 121  
Qy 131 CLDELRGAAAFVVASGSLPPGVADYORVADICRRSSTPLILDTSGGGLQ-HISSG 189  
Db 122 MLAVIQPSNDFEYFVSSGTPKGVPRPFYRLSQITLQKARLLIDTSGAALKHCKEKG 181  
Qy 190 VFLTKASVRELRECVSELTTEPEQLAAHELIDRGAEVVVSLGSGALLATRHASHR 249  
Db 182 IFLCKPNNELSEIVGELITTEKQOEAAAMEIINSKVEITLVVSLGAGAFASKQGIYH 241  
Qy 250 FSSIPMTAVSGVAGADAMVAATVGLSRGWSLIKSVRLGNAAGAMLTPGTAACNRDDV 309  
Db 242 VTAPVREKSTVAGAGDSMVAGWMLSLSRGNHVEDVIKFGVACGTAATMNFETELCKLEDV 301  
Qy 310 ERFFE 314  
Db 302 ERLYK 306

RESULT 6  
US-10-282-122A-50578  
; Sequence 50578, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Onisen, Karl  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; PRIOR FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/151,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23

```

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50578
; LENGTH: 505
; TYPE: PR1
; ORGANISM: Burkholderia mallei
US-10-282-122A-50578
```

```

Query Match      34.7%; Score 597.5; DB 4; Length 505;
Best Local Similarity 41.9%; Pred. No. 2e-47;
Matches 127; Conservative 59; Mismatches 116; Indels 1; Gaps 1;
```

```

QY 12 PRITLTNNPALDITTSVDVVRPTEKMRGAPRYDPPGGGGINVARIYVLCGCSFALFPFA 71
D 194 PEYITLTPNPADIVATCVERVTRTKLRKCPARDPPGGGGINVARIYVLCGCSFALFPFA 253
QY 72 GGSSTSLMALLDGAGVFPFVPIPIASTRESFTVNESRTAKQYRFVLPGBSLTVABEOQC 131
D 254 GCGGTGLALRGLADEGRAHIDLAGETRENFSLERSTGHEPFLVPGPALLAHENPRC 313
QY 132 LDELKGAASAAFFVAGSGLPPGVADYYQYRVADICRSSSTPLILDTSGGGL-HISSGV 190
D 314 VEAIGRLADASRYLVMSGSLPPGMPDDCYAALARAASAGVTVVDTSGPALAALDAGV 373
QY 191 FLKASVRELRECVSESLTEBEOALAAHFLIDGRARVWVSLGSGALLATRHASHRF 250
D 374 YLVPSIGELRALTGLPLEDDGARLAARATVAGGRQIYALITLGDALVSHDDVRL 433
QY 251 SSIPMTAVSGVAGDAMVAATVGLSRGWSLTKSVRLGNAGAAMLLTPGTACNRDVE 310
D 434 PGVAVVARSALIGADSFVAGLVVALNNGANVADAARHAAASASLSTPALTGTKEIDIA 493
QY 311 RPF 313
D 494 RIV 496

RESULT 7
US-10-282-122A-48067
; Sequence 48067, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EUTRA.034A
```

```

; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48067
; LENGTH: 338
; TYPE: PR1
; ORGANISM: Burkholderia cepacia
US-10-282-122A-48067
```

```

Query Match      34.2%; Score 587.5; DB 4; Length 338;
Best Local Similarity 43.0%; Pred. No. 1e-46;
Matches 132; Conservative 48; Mismatches 124; Indels 3; Gaps 3;
```

```

QY 14 IITLTNNPALDITTSVDVVRPTEKMRGAPRYDPPGGGGINVARIYVLCGCSFALFPFA 73
D 31 IYVITLNPADVATVSEHIVDTHTKLCARPRDPGGGGINVARTVIRLGGDCVALLIAGS 90
QY 74 STGSLMALLDGAGVFPFVPIPIASTRESFTVNESRTAKQYRFVLPGBSLTVABEOQCID 133
D 91 PTGDDVLTALLEAHFLPSERIRIAGETRENVCTETATGHEYRPLMGPVLTGEMPRACAA 150
QY 134 ELKGAASAAFFVAGSGLPPGVADYYQYRVADICRSSSTPLILDTSGGGL-HISSGV 192
D 151 RTDAMPSPRYVLTSSSLPPGAPDDLYATLARTAKRGRVWVDAAGRALQALKGVRL 210
QY 193 LKASVRELRECVSESLTEBEOALAAHFLIDGRARVWVSLGSGALLATRHASHRFSS 252
D 211 VKPSIGLSALAG-EPLDITSACRKASSELVANKQADIVALTIGARKAMVTRDRTLRFG 269
QY 253 IPMTAVSGVAGDAMVAATVGLSRGWSLTKSVRLGNAGAAMLLTPGTACNRDVE 312
D 270 RPAAVCSYVAGDPSFVGVWVWALACGVFPDPCRYVLAASASVERGTALTCTRDVERI 329
QY 313 F-ELAE 318
D 330 HAEVLAQ 336

RESULT 8
US-09-815-242-13976
; Sequence 13976, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
```

```

; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13976
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13976

```

Query Match 31.1%; Score 534.5; DB 3; Length 310;

Best Local Similarity 41.4%; Pred. No. 9.3e-42; Matches 126; Conservative 54; Mismatches 119; Indels 5; Gaps 4;

```

Qy 13 RIITLTNNPALDITTSVDVVRPTEKRCGAPRYDGGGINVARIHVHVGCCSTALPPAG 72
Db 3 RIYTLTAPSDSATITTPQIYRBEGLRCSAVFEBGGGINVARAHMLGSTATATIPFAG 62
Qy 73 GSTSLMALALGDAGVPRVPIPIASTRESFTVNESRTAKOYRFLVLPESLTVABEQCL 132
Db 63 GATGEHVSLADENVPVAVTAKDWTRQNLHVHVESGQYRFVMPGATLDDDEFQ-L 121
Qy 133 DELGGAASAFAVVASGLPPGVAADYYQVADICRBSSTPLIIDTSGGGLQHTS--GV 190
Db 122 EEQVLEIESGAILVSSGLPPGVKVEKLTOLISAQKQICITIDSSDALTALAGDI 181
Qy 191 FLKASVRELRECVSSBLTEPEOL-AAAHFLIDRGRAEVVVSIGSGALLATRHASHR 249
Db 182 ELVFPNLKELSAVNRD-LTQPDVRRKAQELVSGKARRVVSIGPGALGISENCIQ 240
Qy 250 FSSIPMTAVSGVAGDAMVAITVGLSRGWSLISVRLGNAAGAMLLTPTAACNRDV 309
Db 241 VVPPVVSQSTVGAGDSWGMATLKLADASLEEMVRFGVAAGSAATLNGTRLCSHDT 300
Qy 310 ERPF 313
Db 301 QKIT 304

```

#### RESULT 9

```

; Sequence 10180, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Hasebeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyckind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242

```

```

; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10180
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10180

```

Query Match 30.9%; Score 531.5; DB 3; Length 309;

Best Local Similarity 41.1%; Pred. No. 1.8e-41; Matches 125; Conservative 55; Mismatches 119; Indels 5; Gaps 4;

```

Qy 13 RIITLTNNPALDITTSVDVVRPTEKRCGAPRYDGGGINVARIHVHVGCCSTALPPAG 72
Db 3 RIYTLTAPSDSATITTPQIYRBEGLRCSAVFEBGGGINVARAHMLGSTATATIPFAG 62
Qy 73 GSTSLMALALGDAGVPRVPIPIASTRESFTVNESRTAKOYRFLVLPESLTVABEQCL 132
Db 63 GATGEHVSLADENVPVAVTAKDWTRQNLHVHVESGQYRFVMPGATLDDDEFQ-L 121
Qy 133 DELGGAASAFAVVASGLPPGVAADYYQVADICRBSSTPLIIDTSGGGLQHTS--GV 190
Db 122 EEQVLEIESGAILVSSGLPPGVLEKLTOLISAQKQICITIDSSDALTALAGDI 181
Qy 191 FLKASVRELRECVSSBLTEPEOL-AAAHFLIDRGRAEVVVSIGSGALLATRHASHR 249
Db 182 ELVFPNLKELSAVNRD-LTQPDVRRKAQELVSGKARRVVSIGPGALGISENCIQ 240
Qy 250 FSSIPMTAVSGVAGDAMVAITVGLSRGWSLISVRLGNAAGAMLLTPTAACNRDV 309
Db 241 VVPPVVSQSTVGAGDSWGMATLKLADASLEEMVRFGVAAGSAATLNGTRLCSHDT 300
Qy 310 ERPF 313
Db 301 QKIT 304

```

#### RESULT 10

```

; Sequence 851, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 851
; LENGTH: 309
; TYPE: PRT

```



```
; ORGANISM: Escherichia coli
US-10-369-493-851
Query Match      30.9%; Score 531.5; DB 4; Length 309;
Best Local Similarity 41.1%; Pred. No. 1.8e-41;
Matches 125; Conservative 55; Mismatches 119; Indels 5; Gaps 4;

Cy 13 RIITLTNNPALDITTSVDVVRPTEKMRGAPRYPDGGGGINVARIHVHVGCCSTALFPAG 72
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 3 RIYTLTAPLSDATITPQIYPEGKLRCTAPVFEPGGGGINVARAIHAGSATATFPAG 62
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Cy 73 GRTGSLMALDGDAGVPRVPIPIASTRESFTVNESRTAKQYRFLPGPSLTVAEQEQL 132
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 63 GATGEHLVSLADENVPATVEAKDWTRQNLHVHVEASGEQYRFVMPGALNDEFRQ-L 121
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Cy 133 DELGGAASAFAVVASGSLPPGVADYYQYRVADICRRSSTPLIDTSGGGLQHIS--GV 190
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 122 EEOVLEIESGAILIVISGSLPPGVKLEKLTQLISAQKQIGIRCIYDSSEALSAALAIQNI 181
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Cy 191 FLKASVREIRECYGSELLTEPEQL-AAAHELIDRGAEVYVVISLGSQGLLATRASHR 249
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 182 ELVYPNOKELISALVNR-LTQPDVVRKAQGEIVNSGKAKRVVSLGPGGALGVDSENCIO 240
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Cy 250 FSSIPMTAVSGVGADAMVAITVGLSRGSLIKSVRLGNAGAAMLTTPGTAACNRDV 309
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 241 VPPPVKSQSTVGAGDSWGAMTTLKLAENASLEEMVRFVGAAGSAATLNGTRLCSHDDT 300
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Cy 310 ERFF 313
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 301 QKTY 304
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 11
US-10-282-122A-56570
; Sequence 56570, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyckind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
```

```
; Remaining Prior Application data removed - See File Wrapper or PAM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin Version 3.1
; SEQ ID NO 56570
; LENGTH: 309
; TYPE: PR
; ORGANISM: Escherichia coli
US-10-282-122A-56570
Query Match      30.9%; Score 531.5; DB 4; Length 309;
Best Local Similarity 41.1%; Pred. No. 1.8e-41;
Matches 125; Conservative 55; Mismatches 119; Indels 5; Gaps 4;

Cy 13 RIITLTNNPALDITTSVDVVRPTEKMRGAPRYPDGGGGINVARIHVHVGCCSTALFPAG 72
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 3 RIYTLTAPLSDATITPQIYPEGKLRCTAPVFEPGGGGINVARAIHAGSATATFPAG 62
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Cy 73 GRTGSLMALDGDAGVPRVPIPIASTRESFTVNESRTAKQYRFLPGPSLTVAEQEQL 132
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 63 GATGEHLVSLADENVPATVEAKDWTRQNLHVHVEASGEQYRFVMPGALNDEFRQ-L 121
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Cy 133 DELGGAASAFAVVASGSLPPGVADYYQYRVADICRRSSTPLIDTSGGGLQHIS--GV 190
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 122 EEOVLEIESGAILIVISGSLPPGVKLEKLTQLISAQKQIGIRCIYDSSEALSAALAIQNI 181
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Cy 191 FLKASVREIRECYGSELLTEPEQL-AAAHELIDRGAEVYVVISLGSQGLLATRASHR 249
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 182 ELVYPNOKELISALVNR-LTQPDVVRKAQGEIVNSGKAKRVVSLGPGGALGVDSENCIO 240
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Cy 250 FSSIPMTAVSGVGADAMVAITVGLSRGSLIKSVRLGNAGAAMLTTPGTAACNRDV 309
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 241 VPPPVKSQSTVGAGDSWGAMTTLKLAENASLEEMVRFVGAAGSAATLNGTRLCSHDDT 300
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Cy 310 ERFF 313
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 301 QKTY 304
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 12
US-10-282-122A-55963
; Sequence 55963, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyckind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
```





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Biocelebration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 06:17:23 ; Search time 8.13514 Seconds  
(without alignments)  
1192.746 Million cell updates/sec

Title: US-10-617-038-22

Perfect score: 11720  
Sequence: 1 MTEPAWDEKPRITITLMN.....TEVGQDQYVHPIVNPASP 339

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 169630 seqs, 2862289 residues

Total number of hits satisfying chosen parameters: 169630

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/PCRT\_NEW\_PUB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	283	16.5	306	US-10-485-517-185	Sequence 185, App
2	168	9.8	488	US-11-087-099-103	Sequence 303, App
3	165.5	9.6	323	US-11-132-864-7	Sequence 7, Appl
4	164.5	9.6	470	US-11-087-099-11203	Sequence 11023, A
5	164.5	9.6	496	US-11-087-099-6042	Sequence 6042, App
6	161	9.4	345	US-11-096-686-11061	Sequence 11061, A
7	145	8.4	473	US-11-087-099-12426	Sequence 12426, A
8	142.5	8.3	480	US-11-087-099-6193	Sequence 6193, App
9	140.5	8.2	480	US-11-087-099-7506	Sequence 7506, App
10	131	7.6	361	US-11-096-568A-6281	Sequence 6281, App
11	131	7.6	369	US-11-096-568A-6280	Sequence 6280, App
12	129.5	7.5	480	US-11-087-099-6825	Sequence 6825, App
13	128.5	7.5	176	US-11-087-099-7668	Sequence 7668, App
14	122	7.1	320	US-10-467-657-3254	Sequence 3254, App
15	121	7.0	302	US-10-793-626-3062	Sequence 3062, App
16	120	7.0	187	US-11-087-099-11101	Sequence 11101, A
17	116	6.7	212	US-11-096-568A-31212	Sequence 6282, App
18	115.5	6.7	378	US-11-096-568A-31212	Sequence 31212, A
19	115.5	6.7	379	US-11-096-568A-31211	Sequence 31211, A
20	115.5	6.7	705	US-11-096-568A-31090	Sequence 31090, A
21	115.5	6.7	741	US-11-096-568A-31089	Sequence 31089, A
22	115.5	6.7	750	US-11-096-568A-31088	Sequence 31088, A
23	114	6.6	397	US-11-132-864-28	Sequence 28, Appl
24	113.5	6.6	428	US-11-096-568A-12725	Sequence 12725, A
25	112	6.5	261	US-11-096-568A-31213	Sequence 31213, A

26	111.5	6.5	316	US-11-087-099-9229	Sequence 9229, App
27	110.5	6.4	477	US-10-520-820-13	Sequence 13, Appl
28	109.5	6.4	323	US-11-087-099-1760	Sequence 1760, App
29	108.5	6.3	307	US-11-087-099-4308	Sequence 4308, App
30	108.5	6.3	310	US-11-087-099-10267	Sequence 10267, A
31	108.5	6.3	323	US-11-096-568A-10266	Sequence 10266, A
32	108.5	6.3	359	US-11-096-568A-10265	Sequence 10265, A
33	106.5	6.2	117	US-09-995-493-40	Sequence 40, Appl
34	106.5	6.2	399	US-11-087-099-2847	Sequence 2847, App
35	105.5	6.1	313	US-11-098-686-11025	Sequence 11025, A
36	105	6.1	319	US-11-087-099-10858	Sequence 10858, A
37	105	6.1	378	US-11-096-568A-12726	Sequence 12726, A
38	101	5.9	319	US-10-793-626-786	Sequence 786, App
39	98.5	5.7	388	US-11-132-864-41	Sequence 41, Appl
40	98.5	5.7	5291	US-11-052-554A-281	Sequence 281, App
41	98	5.7	353	US-11-132-864-29	Sequence 29, Appl
42	98	5.7	380	US-11-132-864-30	Sequence 30, Appl
43	98	5.7	1170	US-11-046-456-28	Sequence 28, Appl
44	98	5.7	1170	US-11-046-644-28	Sequence 28, Appl
45	98	5.7	6893	US-11-205-109-14	Sequence 14, Appl

ALIGNMENTS

```
RESULT 1
US-10-485-517-185
; Sequence 185, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629W0
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 185
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-10-485-517-185

Query Match      16.5%; Score 283; DB 6; Length 306;
Best Local Similarity 26.6%; Pred. No. 2,5e-16;
Matches 81; Conservative 66; Mismatches 139; Indels 18; Gaps 7;

QY      14 IITLVNPAIDITTSVDVVRPTERRCGAPRYPGGGGINVARIVHVLGGCTALPFAAG 73
      2 IYVTFRPSIDVYFTDFKIDGLNRATATYFKAGGGINVSRLKTLVDSTALPAGG 61
      74 STGSLMALLDGACVPRPVPIAASSTREFTVN-ESITAQYQVFLRPSLTVAEQOCL 132
      62 FPGFFIITDNTSNIQSNFIEVDETR---INVKTKTGETEINAGPHITSTQFOQL 117
      133 DELRGAAASAFAVVASGLSPGVAADYORVADICRSSTPLIDTSGGGLQHTSSGV-- 190
      118 QQIKN-TTSDIYIVAGSVSSIPSDAYAOIAQTQAKLVVDAE---KEIAESVLP 172
      191 ---FLKASVRELRCEVSELTPEBQLAAHEDIRGRAEVVVVSGQALLATPHAS 247
      173 YHPLFIKPNDELEWMENTVNSDPTVIKYGRLLVDKG-AQSVIVSGGCAIYIDKEIS 231
      248 HRSSSIIMTVASGVGADANVAIITVGLSRGMSLIKSVRLGNAAGAMLLTPGAACNRD 307
      232 IKAVNPGKVVNTVGSDSITVAGVAGIASGLTIEKAFQAAVACGTAFDEDLA--TRD 289
```

Qy 308 DVER 311  
Db 290 AIEK 293

## RESULT 2

US-11-087-099-303  
; Sequence 303, Application US/11087099  
; Publication No. US20060041961A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53450)B EP  
; CURRENT APPLICATION NUMBER: US/11/087, 099  
; CURRENT FILING DATE: 2005-03-22  
; NUMBER OF SEQ ID NOS: 12464  
; SEQ ID NO 303  
; LENGTH: 488  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-11-087-099-303

Query Match 9.8%; Score 168; DB 7; Length 488;  
Best Local Similarity 22.9%; Pred. No. 1.9e-06;  
Matches 85; Conservative 56; Mismatches 156; Indels 74; Gaps 14;

Qy 11 KPRITITMPALDITTSVDVVRP-----TEKRCGAP--RYDPGGGGINVARIH 59  
Db 83 KPIVSTLGNLCVDIVLSVHSLPPSPGERKALMDELMSPPDKKYWEAGNCMAIAAA 142  
Qy 60 VLGGCSTALPPAGST--GSLMALLDGAGVPRVPIPAASTRESFTYNESRTAKQRPVL 118  
Db 143 RLGHCAVAIGHVGDIEYGEFLDVLHBEGL--GTVALNGTNEKDTSSFCETL--ICWVL 198  
Qy 119 RGP-----SLTAAEBOGL-----DELGAASAAPFVVASGLSPGVADYYQVA 164  
Db 199 VDPQRHGFCSRADPKKEPAPSWITTDLSDEVKMARIOSKVFPCGYPDPDPSPFINSTI 258  
Qy 165 DICRSSTPLILDTSGGGLQHISGV-----FLIKASVRELRCVSGELTPE-- 213  
Db 259 DYAKVGTALFPDGPGRG--KSLSGTDERALLHFLRMSVL-----LITSEVE 308  
Qy 214 -----QLAAHLEIDRGR-AEYVVVSLGSGALLATRHASHRSSIPMTAVSGVAGD 265  
Db 309 ALTGIRPVVAGOEILNKGKGTQWVIVMGPKSILVTKSSVSAVAPAFKEVVDVTYCCGD 368  
Qy 266 AMVAATVVGSRGWSLKSVALGNAAGAMULTPGTA-----ACNRDVER 311  
Db 369 SFVAALALGYIRNPLVNTLTITANAAGATAMGCGAGRNVAKRHQVVDLMKASKLNDDEK 428

Qy 312 FFE-LAAEPT 321  
Db 429 FFEQLLENSE 439

## RESULT 3

US-11-132-864-7  
; Sequence 7, Application US/11132864  
; Publication No. US20050289670A1  
; GENERAL INFORMATION:  
; APPLICANT: Jimui Shi  
; APPLICANT: David Erci  
; APPLICANT: Lisa Hagen  
; APPLICANT: Hongyu Wang  
; TITLE OF INVENTION: Plant Myo-Inositol Kinase  
; FILE REFERENCE: 035718/291638  
; CURRENT APPLICATION NUMBER: US/11/132, 864  
; CURRENT FILING DATE: 2005-05-19  
; PRIOR APPLICATION NUMBER: 60/573, 000  
; PRIOR FILING DATE: 2004-05-20  
; NUMBER OF SEQ ID NOS: 48

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; INFORMATION: Pfam consensus for pfkb family  
US-11-132-864-7

Query Match 9.6%; Score 165.5; DB 7; Length 323;  
Best Local Similarity 28.1%; Pred. No. 1.8e-06;  
Matches 82; Conservative 33; Mismatches 128; Indels 49; Gaps 12;

Qy 47 PGGGGINVARIYVVLGGCSTALPPAGST-----GSLMALLDGAGVPRVPIPAASTRE 101  
Db 35 PGAGANVAVVALRIGPSYKVAFLGKVGDDDFGEFLLELLKKEGVDDYVVDYVDRGTRT 94  
Qy 102 SFTV-----NEARTKQYFVLRPSLTVAE--QEOGLDELGAASAAPFVVASGLSPG 154  
Db 95 GLALVLVDGGERITVYRGA--NADLTLEBDELDLEADITLHSGISLV--LPEP 149  
Qy 155 VAADYQYRVADICRRS-----TPILDTSG--GGLQHISGVFLKASVRELRECV 204  
Db 150 LPBETLEALAAKAGGKISFDPNLRDPLMSDEALEVLELLPLADILKPNBEL-- 205  
Qy 205 GSELT-----EPQLAAHLEIDRGR-AEYVVVSLGSGALLATRHASHRSSIP-MTA 257  
Db 206 --SILTGLKGEVEEALAKLAKAVTKLVVVTGADGALAVTGGGEVHVPAVRKVY 263  
Qy 258 VSGVAGDAMVAALITVGLSR-----GWSLKSVALGNAAGAMULTPG 300  
Db 264 VDTTGAGDAVAGLAGLATDSDNTQDGLDEALRFANMAALVVOCKG 315

## RESULT 4

US-11-087-099-11203  
; Sequence 11203, Application US/11087099  
; Publication No. US20060041961A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53450)B EP  
; CURRENT APPLICATION NUMBER: US/11/087, 099  
; CURRENT FILING DATE: 2005-03-22  
; NUMBER OF SEQ ID NOS: 12464  
; SEQ ID NO 11203  
; LENGTH: 470  
; TYPE: PRT  
; ORGANISM: Oryza sativa (japonica cultivar-group)  
US-11-087-099-11203

Query Match 9.6%; Score 164.5; DB 7; Length 470;  
Best Local Similarity 22.6%; Pred. No. 3.6e-06;  
Matches 84; Conservative 53; Mismatches 181; Indels 53; Gaps 12;

Qy 8 DEKPRITITMPALDITTSVDVVRP-----EKRCGAP--RYDPGGGGINVAR 56  
Db 65 EBKGETDLATLGNLCVDIVLSVPLPPAREEREAYMERLAASPPQKWEAGNCMAIAF 124  
Qy 57 IYVVLG--GSGTALFPAGSTGSLMALLDGAGVPRVPIPAASTRESFTYNESRTAKQYR 115  
Db 125 AAARLIGRCSLGHVBEIYGFLLDVLBEGL--SVGMLNDSGACRWAYETLLCW 182  
Qy 116 FVLPRPSLTVABOECID-----ELRGAASAAPFVVASGLSPGVADYYQYRV 163  
Db 183 LVDPQRHGFCSRADSDERAFSWIKRLPRTETALHNSKILSCNVAPELPRDIYSSA 242  
Qy 164 ADICRSSTPLILDTSGG-----LQIT--SGGVFLKASVRELRECVSEL 208  
Db 243 IDCALDAGTAVFPDPGRKSLHGTLDDEORALEHSLRSLDVLTLTSDAE-----SL 295  
Qy 209 LTEREQLAAHLEIDRG--RAEYVVVSLGSGALLATRHASHRSSIPMTAVSGVAGDAM 267

Db 296 TNNPPIAGGELKRGIRTKWVVKMSKSGSIMVTSKAVSASPSFKIDVDVTVGCCDSF 355  
Qy 268 VAAITVGLSRGMSLTKSVRLGNAAGAMLLPTGTAAACRRDVERFEELAAEPTVEGQDQY 327  
Db 356 TAAIAFGFLHNLPAVSTLTLLANAVGAATATCGG-AGRVVAHLDKVLQLLRE-SNINEDDT 413  
Qy 328 VMHPVNPDEAS 338  
Db 414 PMSCLI--EAS 422

## RESULT 5

US-11-087-099-6042  
; Sequence 6042, Application US/11087099  
; Publication No. US20060041961A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53450)B EP  
; CURRENT APPLICATION NUMBER: US/11/087,099  
; CURRENT FILING DATE: 2005-03-22  
; NUMBER OF SEQ ID NOS: 12464  
; SEQ ID NO 6042  
; LENGTH: 496  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(496)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-11-087-099-6042

Query Match 9.6%; Score 164.5; DB 7; Length 496;  
Best Local Similarity 22.6%; Pred. No. 3.9e-06;  
Matches 84; Conservative 53; Mismatches 181; Indels 53; Gaps 12;

Qy 8 DEGRPIITLTWNPALDITTSVDVVRPT-----EKMRGAP--RYDPGGGGINVAR 56  
Db 87 EEGKETDLATLGNLCVDVLSVLPQLPPAPREREAYMERLAASPPDQKFWEAGNCMLAF 146  
Qy 57 IVHVLG-GCSTALPPAGGSTSLMALLDGAGVFPPIPIAASRESTVNESTRATQYR 115  
Db 147 AARLGRGCTLGHVGEIYKFLDVLLEEGI--SVGMLDNSDSGACRAVAYETLLCWV 204  
Qy 116 FVLPGSLTVAEOECGLD-----ELRGAAASAFAVVASGLPFGVADYQYRV 163  
Db 205 LVDPFGHGFCSGRADFSDPEPAFSWIRKLPATETKAIHHSKTLFCNGIAFDELFPDVISSA 264  
Qy 164 ADICRSSTPLILDTSGGG-----LQHI--SSGVFLKASVRELRECVGSEL 208  
Db 265 IDCAIDAGTAVFPDPGRGKSLMGTLDGQALHSLRLSDVLLTSDAE-----SL 317  
Qy 209 LTEPEQLAAAHLELDRG-RAEVVVSIGSGALLATRHASHRFSSIPMTAVSGVAGADM 267  
Db 318 TNNPPIAGGELKRGIRTKWVVKMSKSGSIMVTSKAVSASPSFKIDVDVTVGCCDSF 377  
Qy 268 VAAITVGLSRGMSLTKSVRLGNAAGAMLLPTGTAAACRRDVERFEELAAEPTVEGQDQY 327  
Db 378 TAAIAFGFLHNLPAVSTLTLLANAVGAATATCGG-AGRVVAHLDKVLQLLRE-SNINEDDT 435  
Qy 328 VMHPVNPDEAS 338  
Db 436 PMSCLI--EAS 444

## RESULT 6

US-11-098-686-11061  
; Sequence 11061, Application US/11098686  
; Publication No. US20060024696A1  
; GENERAL INFORMATION:  
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.  
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES  
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING

; FILE REFERENCE: 09531-128001  
; CURRENT APPLICATION NUMBER: US/11/098,686  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: PCT/US03/31318  
; PRIOR FILING DATE: 2003-10-01  
; PRIOR APPLICATION NUMBER: US 60/416,395  
; PRIOR FILING DATE: 2002-10-04  
; NUMBER OF SEQ ID NOS: 11433  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11061  
; LENGTH: 345  
; TYPE: PRT  
; ORGANISM: Lawsonia intracellularis  
US-11-098-686-11061

Query Match 9.4%; Score 161; DB 7; Length 345;  
Best Local Similarity 21.4%; Pred. No. 4.7e-06;  
Matches 72; Conservative 59; Mismatches 160; Indels 46; Gaps 12;

Qy 11 KPRIT--LTWNPALDITTSVDVVRPTERKRCGAPRYDPGGGGINVARIHVVLGCGSTA 67  
Db 21 KPIVVGDIWLDVNLIGSDRISPEAVPIVKIENKQSLGAG-IVARSIALLGKVTI 79  
Qy 68 LFPAG-GTGSLLMALLDGAGVFPPIPIA-----ASTR-----ESFTVNE 107  
Db 80 IGAVGQDQSGEIKIDLSTRGILSSITTFANRQTVTRVMAHQWIRLDHEESTSYN- 138  
Qy 108 SRTAKOYRFLVPGSLVAEOE--QCDELGAASAFAVVASGLPFGVADYQYRVADI 166  
Db 139 --SKELAMVLSNEKTVSOHEIITLSDYKGLVSKKEFMIG-----FQDIL-L 182  
Qy 167 CRSSSTPLILDTSGGGLQH--ISSGVFLKASVRELRECVGS-ELTEPEQLAAHLELD 223  
Db 183 AKQNPMAVLDPDPNCINMYALCKNIFALPNTKETEGCGMAVTSQPFELAAHGTIMK 242  
Qy 224 RGRAEVVVSIGSGG-ALLATRHASHRFSSIPMTAVSGVAGADMVAITVGLSRGMSLI 282  
Db 243 LLSFKHLITLIGDSGMALFLSPKKIWHIPIVGRDVPDVTGAGGTVIATFGIALSAGIDPL 302  
Qy 283 KSVRLGNAAGAMLLPTGTAAACRRDVERFEELAAEP 319  
Db 303 ISAILANYAAGVAVSQVGTAVSPDELEAITSLPOP 339

## RESULT 7

US-11-087-099-12426  
; Sequence 12426, Application US/11087099  
; Publication No. US20060041961A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53450)B EP  
; CURRENT APPLICATION NUMBER: US/11/087,099  
; CURRENT FILING DATE: 2005-03-22  
; NUMBER OF SEQ ID NOS: 12464  
; SEQ ID NO 12426  
; LENGTH: 473  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(473)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-11-087-099-12426

Query Match 8.4%; Score 145; DB 7; Length 473;  
Best Local Similarity 21.3%; Pred. No. 0.00016;  
Matches 81; Conservative 60; Mismatches 169; Indels 70; Gaps 15;

Qy 4 PAAWDEKPRITLTWNPALDITTSVDVVRPTERKRCG-----PRIDPGGGGI 52  
Db 64 PMSW---KHVDVATLSNLCDIVLVNLPQLPSPPLQKAKFMDRLAASPPDKKYWEAGGNC 120







APPLICANT: MONACI Elisabetha  
TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/10/467,657  
CURRENT FILING DATE: 2003-08-11  
PRIOR APPLICATION NUMBER: GB-0103424.8  
PRIOR FILING DATE: 2001-02-12  
NUMBER OF SEQ ID NOS: 9218  
SOFTWARE: SeqWin99, version 1.04  
SEQ ID NO 3254  
LENGTH: 320  
TYPE: PRT  
ORGANISM: Neisseria gonorrhoeae  
US-10-467-657-3254

Query March 7.1%; Score 122; DB 6; Length 320;  
Best Local Similarity 21.4%; Pred. No. 0.0078;  
Matches 63; Conservative 42; Mismatches 111; Indels 78; Gaps 9;

QY 49 GGGINVARIVHVLGGCTALFPAG-GSTGSLMLLGDAGVPRVPIPAASTRESFTVNE 107  
DB 57 GGAANVARINISLGRAGLISVTGDDAAALDLALMVQDV-----ASYLMRD 104  
QY 108 SR--TAKQYRFLPGPSLTVAEQ-----QCLDELGMAAFAFVVASGLPPGVADY 160  
DB 105 KQITTVVLRVARNQOLRLDPFEPNREVLQIK----- 140  
QY 161 QRVADICRSSTPLILDTSGGGLQHS-----GVFLKA 195  
DB 141 RKYEILPEYVAITFSDYKGGLSHISIMIMAKHVGKTVLIDPKGDYKYGATLITP 200  
QY 196 SVRELRECVSEL---LTPPEOLAAMELIDRGAEVVVSLGSGALLATRHASHRFS 251  
DB 201 NCALKEVVGSKMGEDLTERAQNLRH--LD---LTVALLTRSEBGMTLPSEGEPIYQP 255  
QY 252 SIPMTAVSGVAGDAMVAITVGLSRGWSLIRKSVRLGNAAGAMLLTPGTACN 305  
DB 256 TRAQEVYDVSGAGDVTIAGMGLAAGCTMEAMYLANTAGVYVAKLGTAVCS 309

## RESULT 15

US-10-793-626-3062  
Sequence 3062, Application US/10793626  
Publication No. US20050255478A1  
GENERAL INFORMATION:  
APPLICANT: KIMMERLY, WILLIAM JOHN  
TITLE OF INVENTION: STAPHYLOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
FILE REFERENCE: P03480US  
CURRENT APPLICATION NUMBER: US/10/793,626  
CURRENT FILING DATE: 2004-03-04  
PRIOR APPLICATION NUMBER: 60/164,258  
PRIOR FILING DATE: 1999-11-09  
NUMBER OF SEQ ID NOS: 4472  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3062  
LENGTH: 302  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: synthetic  
OTHER INFORMATION: amino acid sequence  
US-10-793-626-3062

Query March 7.0%; Score 121; DB 6; Length 302;  
Best Local Similarity 23.6%; Pred. No. 0.0087;  
Matches 74; Conservative 54; Mismatches 137; Indels 48; Gaps 15;

QY 26 TTSYD-----VVRPEKMRGAPRYDPGGG-GINVARIVHVLGGCTALFPAG-GSTG 76  
DB 6 STNDKPLNVKRFKPGSTLHINQAKFGGKGNQAIASRLAADTTFISXVKQONA 65  
QY 77 SLMLLGDAGVPRVPIPAAS--TRSF-TVNESRTAKQYRFLPGPSLTVA--EQEQC 131

DB 66 NFLIEDPKKAGIHQVYLITSESEETGQAFITVDEA--GQNTILVYGGAAMTILSATDVENS 123  
QY 132 LDELGMAAFAFVVASGLPPGVADYQVRVADICRSSTPLILDTSGGGLQHSICVF 191  
DB 124 ADAFIG---ADFVVAQLFVP---FEAIRQAFKARKQNTITVLN-----PAPAI 166  
QY 192 LKASVRELRECV-----GSEILT-----EPQOLAAAMELIDRGAEVVVSLGSGAL 240  
DB 167 ELPSLLELTDITIPNETEALLTGISINNESDMKETATYFLDQ-ISAVALITLGEQGY 225  
QY 241 LATRHASHRFSIPMTAVSGVAGDAMVAITVGLSRGWSLIRKSVRLGNAAGAMLLTP 239  
DB 226 CAYQEQYKMLPACNVKALIDTTAAGDTFTIGAFSLKNDLSNISAIRLANQASSLTVQRK 285  
QY 300 GTAAC--NRDDVE 310  
DB 286 GAQASIPTRKEVE 298

Search completed: March 23, 2006, 06:29:31  
Job time : 9.13514 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 04:56:29 ; Search time 181.055 Seconds  
(without alignments)  
1652.628 Million cell updates/sec

Title: US-10-617-038-23

Perfect score: 3508  
Sequence: 1 VLMTAAADYRRSPRVFRD.....LEVTSRWIAGENETPTGL 681

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*  
9: Geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3508	100.0	681	8	AD137302 M. tuberc
2	547.5	15.6	226	9	ABM91252 M. xanthu
3	514	14.7	217	9	ADU49950 Methanosa
4	464	13.2	220	9	AE836491 L. pneumo
5	463	13.2	216	9	AE839892 L. pneumo
6	431	12.3	443	9	AE895281 M. tubercu
7	431	12.3	443	9	AE893950 Mycobacte
8	420.5	12.0	233	9	ADU49942 Meloidogy
9	420.5	12.0	233	9	AE895272 M. javanic
10	420.5	12.0	233	9	AE893956 Root-knot
11	417.5	11.9	229	9	ABM96175 M. xanthu
12	417.5	11.9	233	9	ADU49941 Meloidogy
13	417.5	11.9	233	9	AE895271 M. incogni
14	417.5	11.9	233	9	AE893944 Southern
15	415.5	11.8	289	7	ADB74647 Mycobacte
16	414.5	11.8	257	9	ADU49952 Azotobact
17	413.5	11.8	235	9	ADU49952 Mesorhizo
18	412.5	11.8	226	9	ADU49951 Novosphin
19	406	11.6	236	9	ADU49944 Sinorhizo
20	401.5	11.4	229	9	ADU49943 Heteroder
21	401.5	11.4	229	9	AE895273 H. glycine
22	401.5	11.4	229	9	AE893945 Root-knot
23	401.5	11.4	229	9	AE893946 Soybean c
24	346.5	9.9	227	9	ABM93647 M. xanthu

25	221	6.3	419	1	AA82072	AA82072 Esterase
26	157	4.5	236	2	AAW98370	AAW98370 H. pylori
27	156.5	4.5	4635	8	ADX56098	ADX56098 Streptomy
28	152	4.3	234	2	AAU11078	AAU11078 H. pylori
29	149	4.2	2552	6	ABU33748	ABU33748 Protein e
30	147.5	4.2	1650	7	ABO73052	ABO73052 Pseudomon
31	147	4.2	996	6	ABU22241	ABU22241 Protein e
32	145	4.1	3192	4	AAE10128	AAE10128 Streptomy
33	144	4.1	1741	6	ABU22363	ABU22363 Protein e
34	143.5	4.1	2214	9	ABM96789	ABM96789 M. xanthu
35	143	4.1	836	6	ABU41360	ABU41360 Protein e
36	143	4.1	8360	6	AAE35499	AAE35499 Streptomy
37	141.5	4.0	4799	8	ADX56094	ADX56094 Streptomy
38	141	4.0	4105	5	ABP62762	ABP62762 S. roseos
39	141	4.0	7338	7	ADJ72174	ADJ72174 Streptomy
40	140	4.0	168	8	ADN26380	ADN26380 Bacteri
41	139.5	4.0	2130	8	ADQ74677	ADQ74677 Streptomy
42	138	3.9	1620	8	ADQ74676	ADQ74676 Streptomy
43	136.5	3.9	1340	6	ABU19856	ABU19856 Protein e
44	136	3.9	685	7	ABO77977	ABO77977 Pseudomon
45	136	3.9	846	8	ADN26449	ADN26449 Bacteri

## ALIGNMENTS

RESULT 1	AD137302 standard; protein; 681 AA.
ID	AD137302
XX	AD137302;
AC	AD137302;
XX	22-APR-2004 (first entry)
DT	22-APR-2004 (first entry)
XX	M. tuberculosis low oxygen induced antigen Rv2030c SEQ ID NO:23.
DE	M. tuberculosis low oxygen induced antigen Rv2030c SEQ ID NO:23.
XX	Mycobacterial infection; vaccine; tuberculosis;
KW	Mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;
KM	Mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;
KW	low oxygen induced antigen.
XX	low oxygen induced antigen.
OS	Mycobacterium tuberculosis.
XX	Mycobacterium tuberculosis.
PN	MO2004006952-A2.
XX	MO2004006952-A2.
PD	22-JAN-2004.
XX	22-JAN-2004.
PF	08-JUL-2003; 2003WO-DK000477.
XX	08-JUL-2003; 2003WO-DK000477.
PR	13-JUL-2002; 2002DK-00001098.
XX	13-JUL-2002; 2002DK-00001098.
PA	(STAT-) STATENS SERUM INST.
XX	(STAT-) STATENS SERUM INST.
PI	Andersen P, Rosenkrands I, Stryhn A;
XX	Andersen P, Rosenkrands I, Stryhn A;
DR	WPI; 2004-122778/12.
XX	WPI; 2004-122778/12.
PT	N-PSDB; AD137347.
XX	N-PSDB; AD137347.
PT	Use of one or more polypeptides or their fragments, which are expressed during the latent stage of the mycobacterial infection, and/or nucleic acids encoding the polypeptides, for a therapeutic vaccine against tuberculosis.
XX	Use of one or more polypeptides or their fragments, which are expressed during the latent stage of the mycobacterial infection, and/or nucleic acids encoding the polypeptides, for a therapeutic vaccine against tuberculosis.
PS	Claim 3; SEQ ID NO 23; 76pp; English.
XX	Claim 3; SEQ ID NO 23; 76pp; English.
CC	The present invention describes polypeptides or their fragments, which are expressed during the latent stage of a mycobacterial infection, and/or nucleic acids encoding the polypeptides, which are useful for creating a therapeutic vaccine against tuberculosis. Also described: (1) a therapeutic vaccine against tuberculosis comprising one or more polypeptides; (2) a method for treating an animal, including a human being, with tuberculosis caused by virulent mycobacteria, e.g. by Mycobacterium tuberculosis, M. africanum or M. bovis; (3) a method for immunising an animal, including a human being, against tuberculosis
CC	The present invention describes polypeptides or their fragments, which are expressed during the latent stage of a mycobacterial infection, and/or nucleic acids encoding the polypeptides, which are useful for creating a therapeutic vaccine against tuberculosis. Also described: (1) a therapeutic vaccine against tuberculosis comprising one or more polypeptides; (2) a method for treating an animal, including a human being, with tuberculosis caused by virulent mycobacteria, e.g. by Mycobacterium tuberculosis, M. africanum or M. bovis; (3) a method for immunising an animal, including a human being, against tuberculosis

CC caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis  
CC caused by virulent mycobacteria in an animal, including a human being;  
CC (5) a method for diagnosing previous or ongoing infection with a virulent  
CC mycobacterium; and (6) a method of diagnosing Mycobacterium tuberculosis  
CC infection in a subject. The polypeptides have antibacterial activities,  
CC and can be used in vaccines and in gene therapy. The polypeptides are  
CC useful for the manufacture of a therapeutic vaccine for treating an  
CC individual who is infected by a virulent mycobacterium, e.g. M.  
CC tuberculosis, and who is not vaccinated with BCG against tuberculosis.  
CC The present sequence represents a low oxygen induced antigen, which is  
CC used in the exemplification of the present invention.

SQ Sequence 681 AA;

Query Match 100.0%; Score 3508; DB 8; Length 681;

Best Local Similarity 100.0%; Pred. No. 0;  
Matches 681; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLMTAAADVTRRSRRVRPRDRREGRYLAETLLAAYRDPDVIYGLARGGLPVAMEVAAA 60  
DB 1 VLMTAAADVTRRSRRVRPRDRREGRYLAETLLAAYRDPDVIYGLARGGLPVAMEVAAA 60  
QY 61 LHAFLDAFVVRKLGAPGHDEFAVGALASGGRVNVNDVVRGLRTTPOOLRIAREGREL 120  
DB 61 LHAFLDAFVVRKLGAPGHDEFAVGALASGGRVNVNDVVRGLRTTPOOLRIAREGREL 120  
QY 121 LRRESAYRGERPPTDITGKTVIIVDDGLATGASMFPAVQALRDAQPAQIVAVPAPEST 180  
DB 121 LRRESAYRGERPPTDITGKTVIIVDDGLATGASMFPAVQALRDAQPAQIVAVPAPEST 180  
QY 181 CREPAGLVDDVVCMTMPTPLAVGESFMDPROQVDEYRRLAFTPTGSPSRRRAASTAA 240  
DB 181 CREPAGLVDDVVCMTMPTPLAVGESFMDPROQVDEYRRLAFTPTGSPSRRRAASTAA 240  
QY 241 DVLRRVAIDAPGVPTEHEVLAELVGDARIYVIGESSHGTHEFYQARAAMTOMLIEEGFG 300  
DB 241 DVLRRVAIDAPGVPTEHEVLAELVGDARIYVIGESSHGTHEFYQARAAMTOMLIEEGFG 300  
QY 301 AVAAEADWMPDAYRVNRYVRGLGEDTNADEALSGFERPPAMMMRNTVVRDFVEMLRTRNOR 360  
DB 301 AVAAEADWMPDAYRVNRYVRGLGEDTNADEALSGFERPPAMMMRNTVVRDFVEMLRTRNOR 360  
QY 361 YESGALFQAGFYGLDYSIHRSTIOEVI SYDKVDPRAAARARAYACFDHACADDDGQAYG 420  
DB 361 YESGALFQAGFYGLDYSIHRSTIOEVI SYDKVDPRAAARARAYACFDHACADDDGQAYG 420  
QY 421 FAAAFAFGSGCERAEVQOLVDVQNALAYARQDGLAEDELFFYAQONAQTVRDAEYVYRA 480  
DB 421 FAAAFAFGSGCERAEVQOLVDVQNALAYARQDGLAEDELFFYAQONAQTVRDAEYVYRA 480  
QY 481 MFGSRVTSWMLRDQMAQTLSLTLDRHLDAEPARIYVAHNSHVGDARATEVMADGQ 540  
DB 481 MFGSRVTSWMLRDQMAQTLSLTLDRHLDAEPARIYVAHNSHVGDARATEVMADGQ 540  
QY 541 LITGQIYREKYGDSRSIGSTYGTYYTAASEWGGIQRKAVPALHGSVEELFHQTADS 600  
DB 541 LITGQIYREKYGDSRSIGSTYGTYYTAASEWGGIQRKAVPALHGSVEELFHQTADS 600  
QY 601 FLVSARLSRDAAPLDVVRIGRAIGVYVLPATEROSHYLVHRPADQPDAMTHIDQTALE 660  
DB 601 FLVSARLSRDAAPLDVVRIGRAIGVYVLPATEROSHYLVHRPADQPDAMTHIDQTALE 660  
QY 661 PLEVTSRWINGENPETYPTGL 681  
DB 661 PLEVTSRWINGENPETYPTGL 681

RESULT 2  
ID ABM91252  
XX ABM91252 standard; protein; 226 AA.  
AC  
XX  
XX

DT 02-JUN-2005 (first entry)

XX M. xanthus protein sequence, seq id 10451.

DE Transgenic plant; DNA replication; gene regulation; gene expression.

XX Myxococcus xanthus.

XX US6833447-B1.

FN 21-DEC-2004.

PD 10-JUL-2001; 2001US-00902540.

XX 10-JUL-2000; 2000US-0217883P.

XX (MONS ) MONSANTO TECHNOLOGY LLC.

PA Goldman BS, Hinkle GT, Slater SC, Wiegand RC;

PI WPI; 2005-028716/03.

DR New substantially purified Myxococcus xanthus nucleic acid molecule  
XX encoding a nitrite reductase, useful for determining gene expression,  
PT identifying mutations in a gene of interest, and for constructing  
PT mutations in a gene of interest.

XX Example 2; SEQ ID NO 10451; 25pb; English.

PS The invention relates to a substantially purified nucleic acid molecule  
XX encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a  
XX recombinant DNA construct for expression of a nitrite reductase gene in a  
CC plant cell, and a plant cell comprising the recombinant DNA construct.  
CC The nucleic acid is useful for determining gene expression, identifying  
CC mutations in a gene of interest, and for constructing mutations in a gene  
CC of interest. Sequences given in records for SEQ IDs 9692-16825 represent  
CC a group of 7134 Myxococcus xanthus proteins and peptides. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from USPTO  
XX

SQ Sequence 226 AA;

Query Match 15.6%; Score 547.5; DB 9; Length 226;

Best Local Similarity 51.8%; Pred. No. 8,3e-43;  
Matches 118; Conservative 30; Mismatches 73; Indels 7; Gaps 2;

QY 12 RSPRRVFRDRREGRYLAETLLAAYRDPDVIYGLARGGLPVAMEVAAALHAFLDAFVVR 71  
DB 2 REPE--FQDRYEGGRVLAIDLRRHARQGTIVLALRGGVPGCAEYARKLGVLDVFLVR 59  
QY 72 KLGAPEGHDEFAVGALASGGRVNVNDVVRGLRTTPOOLRIAREGRELRLRESAYRGER 131  
DB 60 KLGTPGHEELAMGALISGGRVLRNREVNELNGRQIEAARREAVELQRRRESYRGER 119  
QY 132 PPTDITGKTVIIVDDGLATGASMFPAVQALRDAQPAQIVYIYVPAPESTCREPAGLVDDV 191  
DB 120 PPTDITGKTVIIVDDGLATGTRAAVVALRLLEPAIIVAVVPAPESTCESLAQVADEV 179  
QY 192 VCATMPTPLAVGESFMDPROQVDEYRRLAFTPTGSPSRRRAASTA 239  
DB 180 ICRWMPPEPYSVGLMYRDFRAQTEDEVRALMEAT-----REGAASEA 222

RESULT 3  
ID ADU49950  
XX ADU49950 standard; protein; 217 AA.  
AC  
XX  
XX  
XX  
XX 27-JAN-2005 (first entry)  
DE Methanosaarcina barkeri ppPT-like protein.  
XX

KW nematocide; purine/pyrimidine phosphoribosyl transferase-like;  
 KM PPPT-like polypeptide; nematode; therapy; enzyme.  
 OS Mehanosarcina barkei.  
 XX  
 XX WO2004096835-A1.  
 PN  
 XX  
 PD 11-NOV-2004.  
 PF 31-MAR-2003; 2003WO-US009532.  
 XX  
 PR 31-MAR-2003; 2003WO-US009532.  
 XX  
 XX (DIVE-) DIVERGENCE INC.  
 PA  
 PI Kloeck AP, Williams DJ, Salmon B,  
 XX  
 XX WPI; 2005-012639/01.  
 DR  
 XX  
 XX Novel purified purine/pyrimidine phosphoribosyl transferase-like  
 PT polypeptide, useful for identifying compound altering activity of  
 PT purine/pyrimidine phosphoribosyl transferase-like polypeptide.  
 XX  
 XX Example; SEQ ID NO 16; 50pp; English.  
 PS  
 XX The invention describes a purified purine/pyrimidine phosphoribosyl  
 CC transferase (PPPT)-like polypeptide (I) comprising an amino acid sequence  
 CC being at least 80% identical to a fully defined sequence (S1) of 233, 233  
 CC or 229 amino acids as given in the specification. Also described are: an  
 CC isolated nucleic acid molecule (II) comprising a nucleotide sequence  
 CC encoding a polypeptide comprising any one of (S1); and an antibody (A1)  
 CC binding specifically to (I); a vector comprising (II); and a transgenic  
 CC cell or transgenic organism having in its genome a transgene containing  
 CC (II). (I) is useful for a method of measuring the binding of a test  
 CC compound to (I), which involves providing (I), contacting a test compound  
 CC to (I), and measuring the binding of the test compound to (I). The method  
 CC further involves measuring PPPT-like activity of (I). The method further  
 CC involves providing a second polypeptide, where the second polypeptide is  
 CC a plant or mammalian PPPT-like polypeptide, contacting the test compound  
 CC to the second polypeptide, and measuring the binding of the test compound  
 CC to the second polypeptide. (I) is useful for identifying a compound that  
 CC alters the activity of (I), which involves providing (I), contacting a  
 CC test compound to (I), and measuring a PPPT-like activity of (I), where a  
 CC change in PPPT-like activity relative to the PPPT-like activity of (I) in  
 CC the absence of the test compound is an indication that the test compound  
 CC alters the activity of (I). The method further involves providing a  
 CC second polypeptide, where the second polypeptide is a plant or mammalian  
 CC PPPT-like polypeptide, contacting the test compound to the second  
 CC polypeptide, and measuring PPPT-like activity of the second polypeptide.  
 CC A1 is useful as therapeutic compound to protect an animal from nematodes,  
 CC as reagents in experimental assays to detect presence of nematodes, as  
 CC tools to screen for expression of the gene product in nematodes, as a  
 CC purification tool of PPPT-like protein, and as PPPT inhibitors/activators  
 CC for therapeutic purposes. This is the amino acid sequence of  
 CC Mehanosarcina barkei PPPT-like protein used in a comparison with  
 CC purine/pyrimidine phosphoribosyltransferase (PPPT) detailed in the  
 CC invention.  
 CC  
 XX  
 SQ Sequence 217 AA;  
 Query Match 14.7%; Score 514; DB 9; Length 217;  
 Best Local Similarity 47.2%; Pred. No. 1.2e-39;  
 Matches 101; Conservative 46; Mismatches 67; Indels 0; Gaps 0;

DB 123 SDSTVILIDGLATGATRAVAISAIRKKNPAKIVAAVPTGAPDTCELFKEVDEVICAT 182  
 Oy 197 PTPFLAVGSFMDPRQVTDDEVRRLATPTAGS 230  
 DB 183 PEPFYGVAMYGNSQTTDEBEVCKLDKAPALRA 216  
 RESULT 4  
 AEB36491  
 ID AEB36491 standard; protein; 220 AA.  
 XX  
 AC AEB36491;  
 XX  
 DT 08-SEP-2005 (first entry)  
 XX  
 DE L. pneumophila protein SEQ ID NO 823.  
 XX  
 KW detection; infection; Antibacterial; Vaccine.  
 XX  
 OS Legionella pneumophila.  
 XX  
 XX WO2005049642-A2.  
 PN  
 XX  
 PD 02-JUN-2005.  
 PF 23-SEP-2004; 2004WO-IB003578.  
 XX  
 PR 21-NOV-2003; 2003FR-00013687.  
 XX  
 PA (INSP) INST PASTEUR.  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 PA (UPLY-) UNIV LYON 1 BERNARD CLAUDE.  
 PA (CNRS) CNRS CENT NAT RECH SCI.  
 XX  
 XX Buchrieser C, Tichit M, Etienne J, Ma L, Cazalec C, Glaeser P;  
 PI Rueschok C, Bouchier C, Zidane N, Magnier A, Kunet F, Vandenesch F;  
 PI Jarraud S;  
 XX  
 DR WPI; 2005-388305/40.  
 XX  
 XX New genome of Legionella pneumophila Paris strain and derived  
 PT polypeptides, useful for detection or identification of the strain and  
 PT for treatment and prevention of infections.  
 XX  
 PS Claim 3; SEQ ID NO 823; 660pp; English.  
 XX  
 XX The invention relates to an isolated or purified nucleotide sequences (I)  
 CC from Legionella pneumophila Paris strain. (I), and their related  
 CC sequences or fragments, are useful as primers and probes for detection  
 CC and amplification, including differentiation between the Paris and  
 CC Philadelphia strains of Legionella pneumophila and to prepare recombinant  
 CC (hybrid) polypeptides (II). (II) are also useful for preparation of  
 CC specific antibodies (Ab), also used for detection/identification of  
 CC Legionella, and some (I), specifically those involved in synthesis of  
 CC surface proteins, are targets for identification of inhibitors. (II), or  
 CC vectors that contain (I), are useful as vaccines and immunogenic  
 CC compositions, for treatment and prevention of infections by L.  
 CC pneumophila. The present sequence represents the amino acid sequence of a  
 CC L. pneumophila protein.  
 CC  
 XX  
 SQ Sequence 220 AA;  
 Query Match 13.2%; Score 464; DB 9; Length 220;  
 Best Local Similarity 45.8%; Pred. No. 6.4e-35;  
 Matches 98; Conservative 40; Mismatches 74; Indels 2; Gaps 1;  
 Oy 9 VTRSRPRVPRDRRGRVLAELAAVARDQPIYVIGARGGLPVAMEVAAALHAPDAF 68  
 DB 1 LSKRMGR--FQDRYQAGQIGATLCKPYANRNDIVIALPRGGGVPVAAAMEBALNAPDLM 58  
 Oy 69 VVRKLGAPGHDEPAVGALASGGRVNVDDVVRGLRTTPQGLRDIABERREGLLRRESAYR 128

```
Db      59  IVRKLSVPGHEBLAFGALAIKIKVNBELLQYIQLSOKDINOVLSEENELARRNAYR 118
Qy      129  GERPPDTITGKTVIVDDGLATGASMPAAVQALRDAQPAQIVIAVPAPESTCFEAGLV 188
      119  GDRKFPDLKKNKIVLIVDDGIATGASMRRAVIALRKESPKKIIIVAVPAPEMDNDIMKSA 178
Qy      189  DDVVCATMPPTPFLAVGESFMDFRQVTDDEVRRL 222
      179  DEVICLISPMSPFNVAWGLMYEDFSQTEDEEVLYLL 212
Db

RESULT 5
AEB39892
ID      AEB39892 standard; protein; 216 AA.
XX
AC      AEB39892;
XX
DT      08-SEP-2005 (first entry)
XX
DE      L. pneumophila protein SEQ ID NO 4224.
XX
KM      detection; infection; Antibacterial; Vaccine.
XX
OS      Legionella pneumophila.
XX
PN      MO2005049642-A2.
XX
PD      02-JUN-2005.
XX
PF      23-SEP-2004; 2004WO-IB003578.
XX
PR      21-NOV-2003; 2003FR-00013687.
XX
PS      (INSP ) INST PASTEUR.
XX
PA      (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PA      (UYLY-) UNIV LYON 1 BERNARD CLAUDE.
XX
PA      (CNRS ) CNRS CENT NAT RECH SCT.
XX
PI      Buchrieser C, Tichit M, Etienne J, Ma L, Cazalet C, Glaser P,
PI      Ruenloek C, Bouchier C, Zidane N, Magnier A, Kunst F, Vandenesch F,
PI      Jarrard S,
XX
DR      WPI; 2005-388305/40.
XX
PT      New genome of Legionella pneumophila Paris strain and derived
PT      polypeptides, useful for detection or identification of the strain and
PT      for treatment and prevention of infections.
XX
PS      Claim 3; SEQ ID NO 4224; 660bp; English.
XX
XX
CC      The invention relates to an isolated or purified nucleotide sequences (I)
CC      from Legionella pneumophila Paris strain. (I), and their related
CC      sequences or fragments, are useful as primers and probes for detection
CC      and amplification, including differentiation between the Paris and
CC      Philadelphia strains of Legionella pneumophila and to prepare recombinant
CC      (hybrid) polypeptides (II). (II) are also useful for preparation of
CC      specific antibodies (Ab), also used for detection/identification of
CC      Legionella, and some (I), specifically those involved in synthesis of
CC      surface proteins, are targets for identification of inhibitors. (II), or
CC      vectors that contain (I), are useful as vaccines and immunogenic
CC      compositions, for treatment and prevention of infections by L.
CC      pneumophila. The present sequence represents the amino acid sequence of a
CC      L. pneumophila protein.
XX
SQ      Sequence 216 AA:
Query Match      13.2%; Score 463; DB 9; Length 216;
Best Local Similarity 46.8%; Pred. No. 7.7e-35;
Matches 96; Conservative 37; Mismatches 72; Indels 0; Gaps 0;
Qy      18  FDRDRRGRVLAELAAVRODPVIVGIARGGLPVAMEVAAALHAPLDAFVVRKLGAPG 77
      4  FDRRYAGQLGAKLAKFYANRNDVIALPRGSPVPAAMISEATNAPLDLMTYRKLGVPG 63
Db
```

```
Qy      78  HDEFVAGALASGRGVVNDVVRGLRITPQQLDIAREBERELLRESAYRGRPPPTDT 137
      64  HEELAFGALAIKIKIKYNEELLQYIQLSODINQVLSQENELARRNRAARGGKPEPDLK 123
Db
Qy      138  GKTIVIVDDGLATGASMPAAVQALRDAQPAQIVIAVPAPESTCFEAGLVDDVVCATMP 197
      124  NKIVLIVDDGIATGASMRRAVIALRKESPKKIIIVAVPAPEMDNDIMKSADEVICLISP 183
Db
Qy      198  TPFLAVGESFMDFRQVTDDEVRRL 222
      184  MSFNAVGLMYEDFSQTEDEEVLYLL 208
Db

RESULT 6
AEB95281
ID      AEB95281 standard; protein; 443 AA.
XX
AC      AEB95281;
XX
DT      06-OCT-2005 (first entry)
XX
DE      M tuberculosis purine/pyrimidine phosphoribosyl transferase - SEQ ID 10.
XX
KM      purine/pyrimidine phosphoribosyl transferase; vaccine; pesticide;
XX      nematode infection; nematocide; enzyme.
XX
OS      Mycobacterium tuberculosis.
XX
PN      US2005172363-A1.
XX
PD      04-AUG-2005.
XX
PF      16-FEB-2005; 2005US-00058869.
XX
PR      30-MAR-2001; 2001US-0280192P.
XX
PR      29-MAR-2002; 2002US-00113201.
XX
PR      21-JAN-2003; 2003US-00348317.
XX
PA      (DIVE-) DIVERGENCE INC.
XX
PI      Kloeck AP, Williams DJ, Salmon B;
XX
DR      WPI; 2005-532486/54.
XX
PT      New isolated purine/pyrimidine phosphoribosyl transferase nucleic acid
PT      and polypeptide, useful as vaccines, pesticides for plants, soil, or
PT      seeds to confer nematode resistance, or drugs treating diseases caused by
PT      nematodes.
XX
PS      Example; SEQ ID NO 10; 21bp; English.
XX
XX
CC      The invention comprises the amino acid and coding sequence of a
CC      purine/pyrimidine phosphoribosyl transferase. The DNA and protein
CC      sequences of the invention are useful as vaccines, pesticides, and drugs
CC      for protecting plants and animals against nematode infection. The present
CC      amino acid sequence represents a Mycobacterium tuberculosis
CC      purine/pyrimidine phosphoribosyl transferase enzyme.
XX
SQ      Sequence 443 AA:
Query Match      12.3%; Score 431; DB 9; Length 443;
Best Local Similarity 32.4%; Pred. No. 2.6e-31;
Matches 129; Conservative 52; Mismatches 161; Indels 56; Gaps 11;
Qy      16  RVRDRRGRVLAELAAVRODPVIVGIARGGLPVAMEVAAALHAPLDAFVVRKLG 75
      2  KLRDDGDKGRQIAQSLAQSGRA-VVVGGLPRGSPVAFVAKXSQAPLDLVVRKLG 60
Db
Qy      76  PGHDEFVAGALASGGGVVNDVVRGLRITPQQLDIAREBERELLRESAYRGRPPPTD 135
      61  PFGPELAFGALISGDGVIVNDVVRGTHIDAAAMDAVERKQLIELRRAERFRGRDRI 120
Db
```

```

QY 136 ITGKTVIVDDGLATGASMPFAVQALRDPAQIVIAVPAPESTCEFAGLVDDVVCAT 195
DB 121 LTGRIAVIVDDGIATGATAKACQVABAGADKVLAIPGPDIVARFAGYADEVVCCLA 180
QY 196 MPTFLAVGESFMDFRQVTDDEEVRRLATPTAGPSLRPAASTADVLRRVAIDAPGCV 255
DB 181 TPALFFAVGGGYRNFTQTSDEVAFL-----DRAHRDPAEAGALIDAAADPP 227
QY 256 THEVLAEVGDARIVLIGESSHGTHEFYQARAAMTQWLIIEKGFQAVA-AEADMPDAYRV 314
DB 228 LRDEEVGVVAGP-----VPVAGHLT-----VPEKPRGIVVFAGSGSSSRHSI 269
QY 315 -NRVY-----RGLGEDTNTADALSGFER-----FPAMMMRNTVVRDPEWMLRTNQRYES 363
DB 270 RNRVVAEVLTGAGFATLLFDLLTPEEERNRANVPDIELLASRLI-DVTGMLATQ----PD 324
QY 364 GALRQAGFYGLDLYSLHRSIQEVSIVLDKVPRAAARA 401
DB 325 TALSIPVGIFG-----ASTGAGALVLAADDPVNVVRA 355

RESULT 7
ID AEB93950 standard; protein; 443 AA.
AC AEB93950;
DT 06-OCT-2005 (first entry)
DE Mycobacterium tuberculosis PPPT-like protein, SEQ ID NO: 10.
XX
XX Screening; antibody production; nematode infection; nematocide;
KW infection; purine/pyrimidine phosphoribosyl transferase-like protein.
XX
XX Mycobacterium tuberculosis.
OS
XX
XX US2005172350-A1.
XX
XX 04-AUG-2005.
XX
XX 27-JAN-2005; 2005US-00045942.
XX
XX 30-MAR-2001; 2001US-0280192P.
XX 29-MAR-2002; 2002US-00113201.
XX 21-JAN-2003; 2003US-00347776.
XX
XX (DIVE-) DIVERGENCE INC.
XX
XX Kloeck AP, Williams DJ, Salmon B;
XX
XX WPI; 2005-563091/57.
XX
XX
XX Screening for inhibitors of nematode purine/pyrimidine phosphoribosyl
PT transferase (PPPT), useful for treating nematode infections (especially
PT in plants), comprises contacting a test compound with the enzyme from
PT Meloidogyne incognita.
XX
XX
XX Example; SEQ ID NO 10; 29pp; English.
XX
XX The invention relates to a method of screening for inhibitors of nematode
CC purine/pyrimidine phosphoribosyl transferase (PPPT, also known as PPRase)
CC which involves contacting a test compound with the enzyme from
CC Meloidogyne incognita. The invention also relates to methods for antibody
CC production. The method is useful for screening for compounds that alter
CC the activity of PPPT. Compounds identified by the method, especially
CC compounds that have a more powerful effect on the nematode enzyme than on
CC plant or mammalian PPPTs, are useful for treating nematode infections,
CC especially in plants. The present sequence is the Mycobacterium
CC tuberculosis PPPT-like protein.
XX
XX Sequence 443 AA;
XX
XX Query Match 12.3%; Score 431; DB 9; Length 443;

```

```

Best Local Similarity 32.4%; Pred. No. 2,6e-31;
Matches 129; Conservative 52; Mismatches 161; Indels 56; Gaps 11;

QY 16 RVFDRREARGVLAELAAAYRDQPDVVLGLAAGLPPVAVEVAALHAPDAPVVRXIGA 75
DB 2 KLPDDRGRDAGRQLAQRLAQLSGKA-VVVLGLPRGGVVAEVAKSLQAPDVLVVRXLGV 60
QY 76 PGHDEFVAGLAAGGRVNVDDVYRGLRITPQQLRDIAREGBELLRRESAYGGERPTD 135
DB 61 PFOPELAFGAIGBDGVAVLNDVVRGTHLDAAAMDVERQQLDELQRRARFRGRDRIP 120
QY 136 ITGKTVIVDDGLATGASMPFAVQALRDPAQIVIAVPAPESTCEFAGLVDDVVCAT 195
DB 121 LTGRIAVIVDDGIATGATAKACQVABAGADKVLAIPGPDIVARFAGYADEVVCCLA 180
QY 196 MPTFLAVGESFMDFRQVTDDEEVRRLATPTAGPSLRPAASTADVLRRVAIDAPGCV 255
DB 181 TPALFFAVGGGYRNFTQTSDEVAFL-----DRAHRDPAEAGALIDAAADPP 227
QY 256 THEVLAEVGDARIVLIGESSHGTHEFYQARAAMTQWLIIEKGFQAVA-AEADMPDAYRV 314
DB 228 LRDEEVGVVAGP-----VPVAGHLT-----VPEKPRGIVVFAGSGSSSRHSI 269
QY 315 -NRVY-----RGLGEDTNTADALSGFER-----FPAMMMRNTVVRDPEWMLRTNQRYES 363
DB 270 RNRVVAEVLTGAGFATLLFDLLTPEEERNRANVPDIELLASRLI-DVTGMLATQ----PD 324
QY 364 GALRQAGFYGLDLYSLHRSIQEVSIVLDKVPRAAARA 401
DB 325 TALSIPVGIFG-----ASTGAGALVLAADDPVNVVRA 355

RESULT 8
ID ADU49942 standard; protein; 233 AA.
AC ADU49942;
DT 27-JAN-2005 (first entry)
DE Meloidogyne javanica purine/pyrimidine phosphoribosyltransferase.
XX
XX nematocide; purine/pyrimidine phosphoribosyl transferase-like;
KW PPPT-like polypeptide; nematode; therapy; enzyme.
XX
XX Meloidogyne javanica.
OS
XX
XX WO2004096835-A1.
XX
XX 11-NOV-2004.
XX
XX 31-MAR-2003; 2003WO-US009532.
XX
XX 31-MAR-2003; 2003WO-US009532.
XX
XX (DIVE-) DIVERGENCE INC.
XX
XX Kloeck AP, Williams DJ, Salmon B;
XX
XX WPI; 2005-012639/01.
XX N-PSDB; ADU49939, ADU49954.
XX
XX Novel purified purine/pyrimidine phosphoribosyl transferase-like
PT polypeptide, useful for identifying compound altering activity of
PT purine/pyrimidine phosphoribosyl transferase-like polypeptide.
XX
XX Claim 4; SEQ ID NO 5; 50pp; English.
XX
XX The invention describes a purified purine/pyrimidine phosphoribosyl
CC transferase (PPPT)-like polypeptide (I) comprising an amino acid sequence
CC being at least 80% identical to a fully defined sequence (S1) of 233, 233
CC or 229 amino acids as given in the specification. Also described are: an
CC isolated nucleic acid molecule (II) comprising a nucleotide sequence

```

CC encoding a polypeptide comprising any one of (SI); and an antibody (A1)  
 CC binding specifically to (I); a vector comprising (II); and a transgenic  
 CC cell or transgenic organism having in its genome a transgene containing  
 CC (II). (I) is useful for a method of measuring the binding of a test  
 CC compound to (I), which involves providing (I), contacting a test compound  
 CC to (I), and measuring the binding of the test compound to (I). The method  
 CC further involves measuring PPPT-like activity of (I). The method further  
 CC involves providing a second polypeptide, where the second polypeptide is  
 CC a plant or mammalian PPPT-like polypeptide, contacting the test compound  
 CC to the second polypeptide, and measuring the binding of the test compound  
 CC to the second polypeptide. (I) is useful for identifying a compound that  
 CC alters the activity of (I), which involves providing (I), contacting a  
 CC test compound to (I), and measuring a PPPT-like activity of (I), where a  
 CC change in PPPT-like activity relative to the PPPT-like activity of (I) in  
 CC the absence of the test compound is an indication that the test compound  
 CC alters the activity of (I). The method further involves providing a  
 CC second polypeptide, where the second polypeptide is a plant or mammalian  
 CC PPPT-like polypeptide, contacting the test compound to the second  
 CC polypeptide, and measuring PPPT-like activity of the second polypeptide.  
 CC A1 is useful as therapeutic compound to protect an animal from nematodes,  
 CC as reagents in experimental assays to detect presence of nematodes, as  
 CC tools to screen for expression of the gene product in nematodes, as a  
 CC purification tool of PPPT-like protein, and as PPPT inhibitors/activators  
 CC for therapeutic purposes. This is the amino acid sequence of Meloidogyne  
 CC Javanica purine/pyrimidine phosphoribosyltransferase (PPPT).  
 CC  
 XX Sequence 233 AA;

SO Query Match

Best Local Similarity 12.0%; Score 420.5; DB 9; Length 233;

Matches 89; Conservative 43; Mismatches 75; Indels 3; Gaps 2;

QY 12 RSPRRVPRDRRAGRVLAELLAAYRDOPD-VIYGLARGGLPVAMEVAAALHAPLDAFVV 70  
 Db 5 RAATAPFKDNDHAGQKLAELAKNFRSQDRKVVLPALPRGVPVAFVAKSLGAPLDLMV 64

QY 71 RKLGAPEHDEFAVGALASGG--RVVNDVVRGLRITPQQLRDIAEBEGRELLRESAYR 128  
 Db 65 RKLGAPEHEBYGIGAVGEGNPPELVNNEADAVKYTPPEGVQAMMEKQLKEIARQRRYL 124

QY 129 GERPPDITGKTVIVDDGLATGASMPFAVQALRDAPQIVIAVPAAPSTCEPAGLV 188  
 Db 125 GDRPPLSLAGKIALIVDDGIGATGTARVAKRLQKNVAKALLASPLAPSDTLAELRAEG 184

QY 189 DDVVCATMPPTPFLAVGESFMDPQVTDDEV 218  
 Db 185 NEVLVLETPPNSAVGLHYTKFDQTSDEEV 214

RESULT 9  
 AEB95272  
 ID AEB95272 standard; protein; 233 AA.

XX AC AEB95272;

XX DT 06-OCT-2005 (first entry)  
 XX DE M javanica purine/pyrimidine phosphoribosyl transferase - SEQ ID 5.  
 XX KM purine/pyrimidine phosphoribosyl transferase; vaccine; pesticide;  
 XX KM nematode infection; nematocide; enzyme.

XX OS Meloidogyne javanica.

XX PN US2005172363-A1.

XX PD 04-AUG-2005.

XX PF 16-FEB-2005; 2005US-00058869.  
 XX PR 30-MAR-2001; 2001US-0280192P.  
 XX PR 29-MAR-2002; 2002US-00113201.  
 XX PR 21-JAN-2003; 2003US-00348317.

XX PA (DIVE-) DIVERGENCE INC.

XX PI Kloeck AP, Williams DJ, Salmon B;

XX DR WPI; 2005-532466/54.

XX DR N-PSDB; AEB95269, AEB95275.

XX PT New isolated purine/pyrimidine phosphoribosyl transferase nucleic acid  
 PT and polypeptide, useful as vaccines, pesticides for plants, soil, or  
 PT seeds to confer nematode resistance, or drugs treating diseases caused by  
 PT nematodes.

XX PS Example; SEQ ID NO 5; 21bp; English.

XX CC The invention comprises the amino acid and coding sequence of a  
 CC purine/pyrimidine phosphoribosyl transferase. The DNA and protein  
 CC sequences of the invention are useful as vaccines, pesticides, and drugs  
 CC for protecting plants and animals against nematode infection. The present  
 CC amino acid sequence represents a Meloidogyne javanica purine/pyrimidine  
 CC phosphoribosyl transferase enzyme.

XX SO Sequence 233 AA;

Query Match 12.0%; Score 420.5; DB 9; Length 233;

Best Local Similarity 42.4%; Pred. No. 9.3e-31;

Matches 89; Conservative 43; Mismatches 75; Indels 3; Gaps 2;

QY 12 RSPRRVPRDRRAGRVLAELLAAYRDOPD-VIYGLARGGLPVAMEVAAALHAPLDAFVV 70  
 Db 5 RAATAPFKDNDHAGQKLAELAKNFRSQDRKVVLPALPRGVPVAFVAKSLGAPLDLMV 64

QY 71 RKLGAPEHDEFAVGALASGG--RVVNDVVRGLRITPQQLRDIAEBEGRELLRESAYR 128  
 Db 65 RKLGAPEHEBYGIGAVGEGNPPELVNNEADAVKYTPPEGVQAMMEKQLKEIARQRRYL 124

QY 129 GERPPDITGKTVIVDDGLATGASMPFAVQALRDAPQIVIAVPAAPSTCEPAGLV 188  
 Db 125 GDRPPLSLAGKIALIVDDGIGATGTARVAKRLQKNVAKALLASPLAPSDTLAELRAEG 184

QY 189 DDVVCATMPPTPFLAVGESFMDPQVTDDEV 218  
 Db 185 NEVLVLETPPNSAVGLHYTKFDQTSDEEV 214

RESULT 10  
 AEB93956  
 ID AEB93956 standard; protein; 233 AA.

XX AC AEB93956;

XX DT 06-OCT-2005 (first entry)  
 XX DE Root-knot nematode PPPT-like protein, SEQ ID NO: 5 #2.  
 XX KM Screening; antibody production; nematode infection; nematocide;  
 XX KM infection; purine/pyrimidine phosphoribosyl transferase-like protein.

XX OS Meloidogyne javanica.

XX PN US2005172350-A1.

XX PD 04-AUG-2005.

XX PF 27-JAN-2005; 2005US-00045942.  
 XX PR 30-MAR-2001; 2001US-0280192P.  
 XX PR 29-MAR-2002; 2002US-00113201.  
 XX PR 21-JAN-2003; 2003US-00347776.

XX PA (DIVE-) DIVERGENCE INC.  
 XX PI Kloeck AP, Williams DJ, Salmon B;









XX 04-AUG-2005.  
 PD  
 XX  
 PF 27-JAN-2005; 2005US-00045942.  
 XX  
 XX 30-MAR-2001; 2001US-0280192P.  
 PR 29-MAR-2002; 2002US-00113201.  
 PR 21-JAN-2003; 2003US-00347776.  
 XX  
 PA (DIVE-) DIVERGENCE INC.  
 XX  
 PI Kloeck AP, Williams DJ, Salmon B;  
 XX  
 DR WPI; 2005-563091/57.  
 DR N-PSDB; AEB93944, AEB93947.  
 XX  
 PT Screening for inhibitors of nematode purine/pyrimidine phosphoribosyl  
 PT transferase (PPPT), useful for treating nematode infections (especially  
 PT in plants), comprises contacting a test compound with the enzyme from  
 PT Meloidogyne incognita.  
 XX  
 PS Claim 14; SEQ ID NO 4; 29pp; English.  
 XX  
 CC The invention relates to a method of screening for inhibitors of nematode  
 CC purine/pyrimidine phosphoribosyl transferase (PPPT; also known as PRTase)  
 CC which involves contacting a test compound with the enzyme from  
 CC Meloidogyne incognita. The invention also relates to methods for antibody  
 CC production. The method is useful for screening for compounds that alter  
 CC the activity of PPPT. Compounds identified by the method, especially  
 CC compounds that have a more powerful effect on the nematode enzyme than on  
 CC plant or mammalian PPPTs, are useful for treating nematode infections,  
 CC especially in plants. The present sequence is the southern root-knot  
 CC nematode PPPT-like protein.  
 CC  
 XX  
 SQ Sequence 233 AA;  
 Query Match 11.9%; Score 417.5; DB 9; Length 233;  
 Best Local Similarity 41.9%; Pred. No. 1.8e-30;  
 Matches 88; Conservative 44; Mismatches 75; Indels 3; Gaps 2;  
 QY 12 RSPRRVFRDRREAGRVIAELIAAYRDQPD-VITGLARGGLPVAMEVAALAHPLDAFVV 70  
 DB 5 RAATAPPKDRHDAQKLAELAKNFKSQRDKVVVLTALPRGVPVAFVAKKSIGADLDLIMV 64  
 QY 71 RKAGPHDEFAVAGALASG--RVVNDVVRGRIITPQQLRDIAEREGRELLRESAYR 128  
 DB 65 RKAGPHDEFAVAGALASG--RVVNDVVRGRIITPQQLRDIAEREGRELLRESAYR 124  
 QY 129 GERPPDTITGKTVIVDDGLATGASMPFAVQALRDAPQAIIVAVPAPESTCREPAGLV 188  
 DB 125 GDRPPLSLAGKIALVDDGIALTGSTARVAMKALQKAVAKALLASPLASPSTLAELEAEG 184  
 QY 189 DDVVCATMPPTPEFLAVGESFMDFRQVTDEEV 218  
 DB 185 NEVLLETPTPFPFSAVGLHYTKFDQTSDEEV 214  
 RESULT 15  
 ADB74647  
 ID ADB74647 standard; protein; 289 AA.  
 AC ADB74647;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Mycobacterium leprae non-naturally occurring peptide #347.  
 XX  
 KM Non-naturally occurring peptide; anion pump protein; tuberculosis;  
 KM hypersensitivity reaction; tuberculostatic.  
 XX  
 OS Mycobacterium leprae.  
 XX  
 PN US6583266-B1.

XX 24-JUN-2003.  
 PD  
 XX  
 PF 16-SEP-1994; 94US-00311731.  
 XX  
 XX 19-AUG-1993; 93US-00109181.  
 PR 22-OCT-1993; 93US-00142558.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Smith DR, Mao J;  
 XX  
 DR WPI; 2003-656441/62.  
 XX  
 PT New Mycobacterium tuberculosis anion pump peptide useful for as  
 PT tuberculosis vaccine and diagnosis of tuberculosis infection.  
 XX  
 PS Disclosure; SEQ ID NO 396; 26pp; English.  
 XX  
 CC The invention relates to a non-naturally occurring peptide of  
 CC Mycobacterium tuberculosis comprising an amino acid sequence  
 CC corresponding to an anion pump protein. The invention also relates to a  
 CC non-naturally occurring nucleic acid corresponding to a DNA sequence of  
 CC Mycobacterium tuberculosis or Mycobacterium leprae. The new peptide is  
 CC useful as a vaccine against Mycobacterium tuberculosis or Mycobacterium  
 CC leprae or for screening for new tuberculosis drugs. Purified proteins  
 CC derived from the sequences of the invention may elicit a specific immune  
 CC response. The peptide may also be used to detect hypersensitivity  
 CC reactions of individuals exposed to Mycobacterium tuberculosis or  
 CC Mycobacterium leprae. The proteins and peptides may be affixed to solid  
 CC supports to detect antibodies typical of hypersensitivity reactions, from  
 CC a patient's sera. This sequence represents Mycobacterium leprae non-  
 CC naturally occurring peptide of the invention. Note: The sequence data for  
 CC this patent did not form part of the printed specification but was  
 CC obtained in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html.  
 CC  
 XX  
 SQ Sequence 289 AA;  
 Query Match 11.8%; Score 415.5; DB 7; Length 289;  
 Best Local Similarity 41.6%; Pred. No. 3.9e-30;  
 Matches 97; Conservative 33; Mismatches 90; Indels 13; Gaps 3;  
 QY 12 RSPRRVFRDRREAGRVIAELIAAYRDQPDVITGLARGGLPVAMEVAALAHPLDAFVV 71  
 DB 53 RNSKRLFDPRVDAGRHLESLRGK-DVVVGLRGSGVPVAFVAKKLRAPLDVLVVR 111  
 QY 72 KLAGPHDEFAVAGALASGGRVVNDVVRGRIITPQQLRDIAEREGRELLRESAYRGER 131  
 DB 112 KLAGPHDEFAVAGALASGGRVVNDVVRGRIITPQQLRDIAEREGRELLRESAYRGER 124  
 QY 132 PPPTDITGKTVIVDDGLATGASMPFAVQALRDAPQAIIVAVPAPESTCREPAGLV 191  
 DB 172 DRAPITGRIVAVDDGIALTGATATKASQVRAQGVSVVLAIVPGRDIFARFAGYADEV 231  
 QY 192 VCATMPPTPEFLAVGESFMDFRQVTDEEVRLR-----ATPTAG-PSLR 232  
 DB 232 VCLHTPAFCVAGGCGCNFTQTSDAEYIALLDARBEFGSATATATGDPQIR 284  
 Search completed: March 23, 2006, 05:10:44  
 Job time : 183.055 secs

This page Blank (uspto)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 05:11:18 ; Search time 27.0938 Seconds  
(without alignments)  
2418.401 Million cell updates/sec

Title: US-10-617-038-23

Perfect score: 3508

Sequence: 1 VLMTAAADVTRSPRRVFRD.....LEVTSRWIAGENPETPTGTL 681

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	3505	99.9	681 2 E70942	hypothetical prote
2	894.5	25.5	658 2 H95852	hypothetical funio
3	780.5	22.2	554 2 T748790	hypothetical prote
4	668	19.0	554 2 C70578	hypothetical prote
5	540	15.4	221 2 AE2055	hypothetical prote
6	431	12.3	443 2 B70933	hypothetical prote
7	276	7.9	554 2 A75296	probable erythromy
8	241	6.9	434 2 F84332	succinoglycan bios
9	224	6.4	419 1 B5E6RM	erythromycin ester
10	220	6.3	175 2 D70305	conserved hypotnet
11	199	5.7	396 2 H67356	hypothetical prote
12	185	5.3	383 2 E75589	hypothetical prote
13	178.5	5.1	387 2 T36697	hypothetical prote
14	159	4.5	446 2 A69750	erythromycin ester
15	157	4.5	240 2 F64671	conserved hypotnet
16	150	4.3	234 2 F71845	hypothetical prote
17	145.5	4.1	621 2 T46851	conserved hypotnet
18	140	4.0	919 2 T37062	probable transcrip
19	136	3.9	574 2 H83117	probable ATP-bindi
20	133.5	3.8	1238 2 T03465	probable exonuclea
21	132	3.8	1185 2 T36761	probable DNA polym
22	131	3.7	1275 2 D38164	cohn protein - pse
23	130.5	3.7	8563 2 T30226	polyketide synthas
24	129.5	3.7	7576 2 T17428	FK506 polyketide s
25	129	3.7	423 2 T11214	3-oxoacyl-lacyl-ca
26	128.5	3.7	537 2 T28683	hypothetical prote
27	128.5	3.7	749 2 A35953	probable bifunctio
28	128.5	3.7	885 2 A84373	leucine-CRNA synth
29	128.5	3.7	3413 2 T17467	rfamycin polyketi

30	128	3.6	7463 2 T36248	CDA peptide synthe
31	127.5	3.6	3164 1 WMBE6	U36 protein - hum
32	127	3.6	384 2 AG2790	conserved hypotnet
33	127	3.6	384 2 E97569	hypothetical prote
34	127	3.6	1361 2 T29435	hypothetical prote
35	127	3.6	1632 2 C70752	probable ctpi prot
36	127	3.6	1937 2 T03224	probable polyketid
37	126.5	3.6	3739 2 T17410	polyketide synthas
38	126	3.6	572 2 T36883	probable DNA repai
39	125.5	3.6	682 2 T36887	probable regulator
40	125	3.6	1116 2 AC2921	cobalamin biosynth
41	125	3.6	1116 2 D97695	cohn protein homol
42	125	3.6	1321 2 T10929	3C3.20c protein -
43	125	3.6	2117 2 T30283	CDA peptide synthe
44	124.5	3.5	6420 2 T30283	polyketide synthas
45	124	3.5	502 2 T50754	protochlorophyllid

ALIGNMENTS

RESULT 1				
E70942				
hypothetical protein Rv2030c - Mycobacterium tuberculosis (strain H37Rv)				
C:Species: Mycobacterium tuberculosis				
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 31-Dec-2004				
C:Accession: E70942				
R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998				
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.				
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome				
A:Reference number: A70500; PMID:98295987; PMID:9634230				
A:Accession: E70942				
A:Status: preliminary; nucleic acid sequence not shown; translation not shown				
A:Molecule type: DNA				
A:Residues: 1-681 <COL>				
A:Cross-references: UNIPROT:O53475; UNIPARC:UPI00000D108D; GB:AL021899; GB:AL123456; NID				
A:Experimental source: strain H37Rv				
C:GeneticS:				
A:Gene: Rv2030c				
C:Superfamily: uncharacterized conserved protein with erythromycin esterase domain				
Query Match				
Best Local Similarity 99.9%; Score 3505; DB 2; Length 681;				
Matches 680; Conservative 1; Mismatches 0; Indels 0; Gaps 0;				
QY	1	VLMTAAADVTRSPRRVFRDREAGRVLAELAAVRDQPVIVGLARGGLPVAMEVAAA	60	
DB	1	MLMTAAADVTRSPRRVFRDREAGRVLAELAAVRDQPVIVGLARGGLPVAMEVAAA	60	
QY	61	LHAPLDAFVVRKUGAPGHDFPAVGALLASGGRVYVNDVVRGRTTPQQLDIAREBEREL	120	
DB	61	LHAPLDAFVVRKUGAPGHDFPAVGALLASGGRVYVNDVVRGRTTPQQLDIAREBEREL	120	
QY	121	LRSSAAYRGERPPTDINGKTVIVVDGLATGASFAAVALRDPQPIYIAVPAABEST	180	
DB	121	LRSSAAYRGERPPTDINGKTVIVVDGLATGASFAAVALRDPQPIYIAVPAABEST	180	
QY	181	CREPAGLVDDVVCATMPETPLAVGESFMDFRQVYDDEVRRLATPTGAPSLRRPAASTAA	240	
DB	181	CREPAGLVDDVVCATMPETPLAVGESFMDFRQVYDDEVRRLATPTGAPSLRRPAASTAA	240	
QY	241	DVLRRAVDAIPGVPYTHETVLAELVGARITVLIGSSSGTHIEFYOARAAMTQMLEEGFG	300	
DB	241	DVLRRAVDAIPGVPYTHETVLAELVGARITVLIGSSSGTHIEFYOARAAMTQMLEEGFG	300	
QY	301	AVAAEADMPDAPYRVRNRVVRGLGEDTNADEALSGPERFPAMMMRTVVRDPEVEMLTRNOR	360	
DB	301	AVAAEADMPDAPYRVRNRVVRGLGEDTNADEALSGPERFPAMMMRTVVRDPEVEMLTRNOR	360	
QY	361	YESGALRQAGFYGLDLYSLHRSIOEVI SYLDKVDPRAAARARAYACFDHACDGOAYG	420	

```
Dh 361 YESALRKQAGFYGLDLSLHRSIOEVIISYLDKVDPRRAARARACFDHACADGQAYG 420
Qy 421 FAAAFGAGPSCEREAVEQLVDVORNALAYABODGLADELFYAQONNAQTVRDAEYVYRA 480
Db 421 FAAAFGAGPSCEREAVEQLVDVORNALAYABODGLADELFYAQONNAQTVRDAEYVYRA 480
Qy 481 MFSGRVTSWMLRDQHMMAQTGSLTLTHDRHLDAPPARIVVAHNSHVGDAATEVMADGQ 540
Db 481 MFSGRVTSWMLRDQHMMAQTGSLTLTHDRHLDAPPARIVVAHNSHVGDAATEVMADGQ 540
Qy 541 LTTGQIVRERYGDSRSIGFSTYGTATYAASEMGIAQRKAVRPAHSGVVELFHQTADS 600
Db 541 LTTGQIVRERYGDSRSIGFSTYGTATYAASEMGIAQRKAVRPAHSGVVELFHQTADS 600
Qy 601 FLVBARLSRDAEAPLDVRLGRAIGVYLPATERQSHLHVRPADQDFAMHIDQTRALE 660
Db 601 FLVBARLSRDAEAPLDVRLGRAIGVYLPATERQSHLHVRPADQDFAMHIDQTRALE 660
Qy 661 PLEVTSRWIAGENPEYPTGL 681
Db 661 PLEVTSRWIAGENPEYPTGL 681
```

## RESULT 2

```
H95852
hypothetical fusion protein [imported] - Sinorhizobium meliloti (strain 1021) megaplaasm1
C/Species: Sinorhizobium meliloti
C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 31-Dec-2004
C/Accession: H95852
R/Finan, T.M.; Weidner, S.; Wong, K.; Buhmeester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A/Title: The complete sequence of the 1,683-kb pSymB megaplaasmid from the N2-fixing endo
A/Reference number: A95842; MUID:21396508; PMID:11481431
A/Accession: H95852
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-658 <KIR>
A/Cross-references: UNIPROT:Q92X74; UNIPARC:UPI00000CB3E6; GB:AL591985; PIDN:CAC48488.1;
R/Experimental source: strain 1021, megaplaasmid pSymB
R/Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abolde, P.; Ampe, F.; Barloy-Hubler,
pel, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A/Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A/Reference number: A96039; MUID:21368234; PMID:11474104
A/Contents: annotation
C/Genetics:
A/Gene: SMD20088
A/Genome: Plasmid
C/Superfamily: uncharacterized conserved protein with erythromycin esterase domain
```

```
Query Match 25.5%; Score 894.5; DB 2; Length 658;
Best Local Similarity 36.7%; Pred. No. 8.2e-50;
Matches 255; Conservative 81; Mismatches 269; Indels 89; Gaps 23;
Qy 36 RDGQDVIVLGIARGLVAVMVAALAPIDAFYVRKLGAGHDEFAV--GALASGRV 93
Db 6 RARRMTTSHLSRRGIDRHHVEMGVPREAFV---DPGFEEFVDSALSIHQGT 60
Qy 94 VNDVNVGLRITPQOLR-----DIAREG--RELARRESAV-----RGRPP 133
Db 61 ISQPYIYALMIRAEVPGDTVEIGSGSYAAVLSRIAAYVTIRHAGLAEVAARRF 120
Qy 134 TDTT-GKTVIVVDGLA--TGASMPAAV-----QALRDAQPAQIVIAVPADEST 180
Db 121 AELRYGNIDRVAGDGPAPGPDALIVAAAGPEIIPHALKEQDLDGSHLIVPGPPE 180
Qy 181 CREPAGLVDDVVCATMTPTPLA-----VGSFMDRQVTVDEVKRLALTPTGSPSLR 233
Db 181 QRLMK--VTRVNAITTFEEDQDGVGVFVPLVGEYGM-----HEBRAQSRAAPRPAPTLPE 232
```

```
Qy 234 PAASTADVLRRAVIDAPGVPTHEVTLAEVLGDAKITVLTSESSHGTFEFQARAAQTQL 293
Db 233 LVAE-AAEPPLBDL-----DDPAFGLFDRFAGRRITVLGEASHGTFEFARAPALITRL 285
Qy 294 IEEKFGAVAAEADMPDAYVNRVYRGLGEDTNADEALSGFERPPAMMNTVVRDVEV 353
Db 286 IEEHGFALVAVEADMPDPAALVDRYVRRRQGITLD--AFQRFPTMMNRREYMDVEV 342
Qy 354 LRTNCRYESGALROAGFYGLDLSLHRSIOEVIISYLDKVDPRRAARARAYACFDHACA 413
Db 343 MREHNGR--KSLDPRAFGFYGLDLYNNRGISALVARYDEIDPEAAVAARRYGLT-PMQ 399
Qy 414 DDGQAGFAAFAAGPSCEREAVEQLVDVORNALAYABODGLADELFYAQONNAQTVR 473
Db 400 NESTYGRAAMTAGFRKCEEAVALVRCQCELEKOLEAGRDGSG---DELDVAONARLYAS 455
Qy 474 AEVYRAMFSGRVTSMRLPDQHMMAQTGSLTLTHDRHLDAPPARIVVAHNSHVGDAAT 533
Db 456 AERYRIMYYAGADSMNMNRDTHMPEITLHLL-----NARGARKAVVAHNSHIGDARHT 510
Qy 534 EV-WADQQLTLAGQIVREYGDERSIGFSTYGTATYAASEMGIAQRKAVRPAHSGVVEE 592
Db 511 DMGARRELINIGQLCRERFEDQALIGFTHGHVAAASDMNDMEVKEIRPSELGSEYER 570
Qy 593 LPHQT-ADSFVLS---ARLSRDAEAPLDVRLGRAIGVYLPATERQSHYHVRPADQF 647
Db 571 VMHDSGVGCEFLIDPARHARLDGLTKPL-----LERFGIVRDPDTERFHYHAASLPQCF 626
Qy 648 DAMIHIDQTRALEPLVTSRWIAGENPEYPTGL 681
Db 627 DGFVWFDQTPVSEPL--GAQHIRAGVDPDTPFGL 658
```

## RESULT 3

```
T48790
hypothetical protein J3E11.380 [imported] - Neurospora crassa
C/Species: Neurospora crassa
C/Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 31-Dec-2004
C/Accession: T48790
R/Schulte, U.; Aign, V.; Hohenstein, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyekatura,
submitted to the Protein Sequence Database, April 2000
A/Reference number: Z24541
A/Accession: T48790
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-554 <SCH>
A/Cross-references: UNIPROT:Q9P6X8; UNIPARC:UPI000017B43F; EMBL:AL535820; GSSDB:GN00112;
A/Experimental source: cosmid contig J3E11; strain 74
C/Genetics:
A/Gene: NCSP:J3E11.380
A/Map position: 2
A/Intons: 346/1
C/Superfamily: uncharacterized conserved protein with erythromycin esterase domain
```

```
Query Match 22.2%; Score 780.5; DB 2; Length 554;
Best Local Similarity 38.8%; Pred. No. 1.4e-42;
Matches 192; Conservative 72; Mismatches 174; Indels 57; Gaps 17;
Qy 219 RLLATPTAGP--SLR-----PAASTADVLRRAVIDAPGVPTHEVLA--ELVGDAR 259
Db 85 ROSTPQPPASPLRLRHHVHPFSLSEALD-----DPKQDIFRYFDSFSSAKV 132
Qy 270 VLGESHGTHGFYQARAAMTQWLIEKGFGAVALAEADMPDAYVNRVYR---GLGE--- 323
Db 133 LLLGDASHGISEYARPAALIRYMIHSHSNITYACEADMPDASIDRYVRRRGPFRAT 132
Qy 324 -DITNADEALSG---PERPPAMMNTVVRDVEVEMLRTNRQYESGALROAGFYGLDYS 378
Db 193 ISEAEARBRAGRPSPFRPMWRMKETHDFEWLRCKNMKGDM--KKEATGFYGLDLYS 251
Qy 379 LHRSGIOEVIISYLDKVDPRRAARARAYACFDHAC--ADGQAYGFAAFAAG--PSGER 433
Db 379 LHRSGIOEVIISYLDKVDPRRAARARAYACFDHAC--ADGQAYGFAAFAAG--PSGER 433
```

Db	252	MGASBRANVIAALBEVDKOMAEATAK---RCYBRILDMSDPDHEHYGLTILVSGKGFCKYEE	308
Qy	434	EAVEQDVNDVQNNALAY--ARODGLAEDELTYAQQNAQTVDAEYTYRPAFMSGRTYSNNL	491
Db	309	DVVOVLTQWLILSKRIEYERAKWDG---BEFHGEQNNALVDADEHYTYAPAMYADESNL	364
Qy	492	RDOHMAQRTGSLTLTHLDRLDAPARIYVMAHNSHVGARATEV-MADGQTLTGQIVVER	550
Db	365	RDTHMFETLQGIIDHRGSH-----SKATVMAHNSHIGARATSMSSWSDDELNVQGLCEA	419
Qy	551	YGDSSRTSGFSTTYTGTVTVAASEWGSIAQRKAVRPALHGSVEELFHQTR-DSFLVLSARISR	609
Db	420	YGKALATGCGYTGKVAARKWDGMSVMHVKPALPNSYSERLMDITGIKNFVLDLRKX	479
Qy	610	DAEAPLDVW---RLGRATGVVYLPAATERQSHYLNHRPAPDOCDAMIHIDQTRALEPLEVTS	666
Db	480	CDKALRELALVEKRLERFIQVITYAPATERQSHSPFVLPBQLDGLAIPDETBEVQALAEVHQ	539
Qy	667	RWINGENPEYTPGL	681
Db	540	PPGPLEFDETWPFGL	554

RESULT 4

C70578

hypothetical protein Rv2143 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004

C:Accession: C70578

R:Colo, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Saltsdon, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: C70578

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-352 <CDL>

A:Cross-references: UNIPROT:O06232; UNIPARC:UPI00000D0FBE; GB:Z95388; GB:AL123456; NID:G

C:Genetics:

A:Experimental source: strain H37RV

A:Gene: Rv2143

Query Match 19.0%; Score 668; DB 2; Length 352;

Best Local Similarity 61.1%; Pred. No. 1.3e-35;

Matches 138; Conservative 25; Mismatches 63; Indels 0; Gaps 0;

QY 3 MTAADADTRESRPVFRDRREAGRVLAELLAAYRDPQVYVIGLARGGLPVAWEVAALH 62

DB 124 MNPSGFLRKPTPTFRFRDRDAGRVLADELASYGRDMLVLGLARGGVPGWEVAASLG 183

QY 63 APPLAFVVRKKGACGHDEFAVGALASGGRVVVNDPVVNGRLRTPOQLDIEREGRELLR 122

DB 184 AELDVFLVRKGVQWRKRLAMGALASGGGVVNDVVSLLITDQGVPAALDSTAEALQR 243

QY 123 RESAVRGERPPTDITGKTVIVDDGLATGASMPAAVQALRDAQPAQIVIAVPAPESTCR 182

DB 244 RELAVRGRGPVVDPRARIVILVDDGIATGASMLAAVFTIRATGESIVAVVPATACR 303

QY 183 EFAGLVDPVVCATMTPEPLAVGESFMDPROVTDSEVRLLATPPAG 228

DB 304 ELAAEADVVCATMPAEFAEAVGQVNDPFOVTDSEVRLLATPTTG 349

RESULT 5

AE2055

hypothetical protein ali1995 [imported] - Noctoc sp. (strain PCC 7120)

C:Species: Noctoc sp. PCC 7120

A:Note: Noctoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

A:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004

A:Accession: AE2055

R.Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Itiguchi, N.; Shimojo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-211, 2001  
A.Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena*  
A.Reference number: AB1807, MUID:21595285; PMID:11759840  
A.Accession: AE2055  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-221 <KUS>  
A.Cross-references: UNIPROT:O8YVJ3; UNIPARC:UPI00000CE2A1; GB:BA000019; PIDN:BA873694.1;  
A.Experimental source: strain PCC 7120  
C.Genetics:  
A.Gene: all1995

Query Match	15.4%	Score 540	DB 2	Length 221
Beet Local Similarity	52.4%	Pred. No. 1.2e-27		
Matches 108	Conservative 41	Mismatches 57	Indels 0	Gaps 0
QY	17	VFRRRREAGRVLAELLAAAYRDQPDVIVIGIARGGLPVAMEVAALALHPIDAFAVVRKLGAP	76	
DB	2	LFKDDRTVAGQVYVAKKLADYANRSNVLTALPRGVPVGFVFAVALNALPLVLVVRKLGVP	61	
QY	77	GHDEFAYCALAGSGRVVNVNDVVRGRITRPQQLRDIAEREGRELLRESAYRGSRPPTDI	136	
DB	62	DNEELAMGAIAGGVRILMQSIQISDEVIAARVAQDERELERRESMTGRGRRPPLL	121	
QY	137	TGKTVIVVDDGLATGASMPAAVOALRDAQPAQIVIVNPAPESTCREFAGIYDDVVCATM	196	
DB	122	KGQTVILVDDGLATGATMMAALIIAVRQQQKEIIVAVPVAAPETCEMQSQVKEKIFCANT	181	
QY	197	PTPEPLAVGESFMDPROVTDSEVRRL	222	
DB	182	PSPEYSVGMMVKEFPQITDDEVRELL	207	

RESULT 6  
 B70933  
 Hypothetical protein RV0571c - Mycobacterium tuberculosis (strain H37Rv)  
 C1Species: Mycobacterium tuberculosis  
 C1Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
 C1Accession: B70933  
 R1Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
 R1Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
 R1Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A1Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A1Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A1Reference number: A70500; MUID:98295987; PMID:9634230  
 A1Accession: B70933  
 A1Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A1Molecule type: DNA  
 A1Residues: 1-443 <CON>  
 A1Cross-References: UNIPROT:O53768; UNIPARC:UPF0000165237; GB:AL021942; GB:AL123456; NID  
 A1Experimental source: strain H37Rv  
 C1Genetics:  
 A1Gene: RV0571c

[illegible]











Query Match 4.5%; Score 159; DB 2; Length 446;  
 Best Local Similarity 21.1%; Pred. No. 0.0092;  
 Matches 92; Conservative 73; Mismatches 174; Indels 96; Gaps 20;

QY 260 LAELVDGARIVLIGSSSGHTEFYQAAATQWLEEGFGVAADPADRVNRVYR 319  
 Db LKTVLKGRIYVLGSGTTHGAGINATKVMIKYHLEHGYDLARESGFP----- 108  
 QY 320 GLGEBTNDDELSCFERPPAMWRTV-----VRPFVEMLRFRNRYSG-ALRQAGFYG 373  
 Db 109 ----DTNA-SYLNMDQLTPKSTMKNSIYAVWHTEDVELFDYMKQKKGDPILITGF-- 161  
 QY 374 LDLYSLHRSIQ-EVYSYLDKVDPRAAARARAYACFDHACADDDQAGYFAAFGAGPSCE 432  
 Db 162 -DIQMKNSFNVAALQWKAADPEKELLSSGENDFTLVDSNTFDEFS----- 210  
 QY 433 REAVEQLVDVQRNALAYAR-----QDGLAEDELFYAQONQTVR--DAEYYRAMESG 484  
 Db 211 -QKKEKLVKNYQKLIKFTKTHASELKENLPKEPKAYEMFMSLQLRIDVETMYLSEKKE 269  
 QY 485 RVT-----SNVLBDQHNQTLGSLITHDLRHLDPARIVVMANSHV--GDARAT 533  
 Db 270 KLEEYPERNIEDFSFMRDRMAAEQFQWVADTL----YPKKCIYWGHNHRLRKQNTYMI 324  
 QY 534 EVMADGOL---TIGQIVRERYGDESRISIGFSTYGTVTAASEMGSIQAKKAVRPAL----- 586  
 Db 325 KQWV--QLNGPFWGDLPERLKKQYTYTIGIYVSGASLDSGSD-----NKTVPKPTSPPP 376  
 QY 587 HGSVEELFHQTADSFVLSARLSRDAEAP---LDVRLGRAIGV--VYLPAERQSHYLV 641  
 Db 377 SGLLEALL-----KADRPVAVFDPLHTKNKKGTSWMTPTPTALYWGVMEE 422  
 QY 642 RP--ADQPDAMIHD 654  
 Db 423 QMILKEQYDGVYIWE 437

RESULT 15

F64671 conserved hypothetical protein HP1214 - Helicobacter pylori (strain 26695)  
 C/Species: Helicobacter pylori  
 C/Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 09-Jul-2004  
 C/Accession: F64671  
 R/Tomb, J.F.; White, O.; Kestavange, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.  
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khairak, H.G.; Glodek, A.; McKen-  
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.  
 Nature 388, 539-547, 1997  
 A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Frazer, C.  
 A/Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
 A/Reference number: A64520; MUID:97394467; PMID:9252185  
 A/Accession: F64671  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-240 <TOM>  
 A/Cross-references: UNIPROT:O25813; UNIPARC:UP100000C0903; GB:AE000627; GB:AE000511; NID  
 C/Genetics:  
 A/Start codon: GTG

Query Match 4.5%; Score 157; DB 2; Length 240;  
 Best Local Similarity 21.5%; Pred. No. 0.0054;  
 Matches 50; Conservative 49; Mismatches 132; Indels 2; Gaps 2;

QY 2 LMTAADVTRRSRPRVFRDRRAGRVLAELLAAYRDQPDVIVLGLARGLPVAMEVAAL 61  
 Db 7 LNTDSHTIDIEGKMFINEEDALNKLINEIHTRHIDLKDSITLALSFNALYLAHALAQKF 66  
 QY 62 HAPLDAFYVKLGAGHDEFAVGALASGRVYVNDVVRGLRITPQQLRDIAEREGRE-L 120  
 Db 67 GATYDILFLPLALNKSCEIALVSESMDIWNESLINSFDITLDYVYGEAKRAYEEDI 126  
 QY 121 LRRESAYGERPPTDITGKTVIVDDGLATGASMPAAVQALRDAQPAQIVIAVPAPEST 180

Db 127 LSHIYQYRKGNAIKSLDKNIFIVDRGIETGFRAGLGVQTCLEKCODIYILTPVAQNV 186  
 QY 181 CREFAGLVYDVVCAATMPTPLAVGESFWDPRQVDEEVRRLATPTAGPSLR 233  
 Db 187 AQGLSCLDGVISYVRPECFVSEHHTYKELRLSNEVEKYLGAANN-PNLKK 238

Search completed: March 23, 2006, 05:26:30  
 Job time : 29.0938 secs

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 04:56:43 ; Search time 174.174 Seconds

(without alignments)  
2758.529 Million cell updates/sec

Title: US-10-617-038-23

Perfect score: 3508

Sequence: 1 VLMTAAADVTRSPRRVFRD.....LEVTSRWIAGENPTPTGL 681

Scoring table: BIOSUM62

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot\_05.80.\*

1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3505	99.9	681	053475_MYCTU	053475 mycobacteri
2	3022	86.1	606	077241_MYCBO	077241 mycobacteri
3	1974	56.1	390	050822_MYCTU	050822 mycobacteri
4	1537	43.8	385	050819_MYCTU	050819 mycobacteri
5	1160.5	33.1	495	05NMQ1_AZOSR	05NMQ1 azoarcus sp
6	1120.5	31.9	448	04H703_SDEIO	04H703 deynococcus
7	956.5	27.3	679	089197_BRHJA	089197 bradyrhizob
8	919	26.2	653	06W168_RHISM	06W168 rhizobium s
9	894.5	25.5	658	092X74_RHIME	092X74 rhizobium m
10	848	24.2	456	06MCX4_PARIW	06MCX4 paracitlmyd
11	840.5	24.0	455	068HX2_PSEBP	068HX2 pseudomonas
12	816	23.3	445	041WA2_AZOV1	041WA2 azotobacter
13	816	23.3	453	05B585_EMENT	05B585 aspergillus
14	811.5	23.1	463	04WJ15_ASPFU	04WJ15 aspergillus
15	800.5	22.8	455	041713_GIBZE	041713 gibberella
16	780.5	22.2	474	09P6X8_NEUCR	09P6X8 neurospora
17	780.5	22.2	554	07RUH3_NEUCR	07RUH3 neurospora
18	773.5	22.0	443	04NE69_9MICC	04NE69 arthrobacte
19	710.5	20.3	428	06MM92_BDEBA	06MM92 bdellovibri
20	678	19.3	352	073YR8_MYCPA	073YR8 mycobacteri
21	668	19.0	352	07TYZ6_MYCBO	07TYZ6 mycobacteri
22	668	19.0	352	06G232_MYCTU	06G232 mycobacteri
23	540	15.4	221	08YV13_ANASP	08YV13 anabaena sp
24	538	15.3	221	08PW22_METMA	08PW22 methanococ
25	537	15.3	235	063X22_BURPS	063X22 burkholderi
26	537	15.3	220	04H704_SDEIO	04H704 deynococcus
27	527	15.0	230	088HX3_PSEBP	088HX3 pseudomonas
28	520	14.8	241	07NLI5_GLOVI	07NLI5 gloeobacter
29	483	13.8	215	041MA5_9BURK	041MA5 burkholderi
30	479	13.7	218	083E61_COXBU	083E61 coxiella bu
31	471	13.4	216	05WXC9_LEGFL	05WXC9 legionella

32	464	13.2	220	2	05ZWC4_LEGPH	05ZWC4 legionella
33	464	13.2	228	2	041YW0_AZOV1	041YW0 azotobacter
34	463	13.2	216	2	05X607_LEGPA	05X607 legionella
35	431	12.3	443	2	053768_MYCTU	053768 mycobacteri
36	431	12.3	443	2	07UIR3_MYCBO	07UIR3 mycobacteri
37	431	12.3	451	2	07DPW2_MYCTU	07DPW2 mycobacteri
38	423	12.1	441	2	06MGV6_BDEBA	06MGV6 bdellovibri
39	418.5	11.9	206	2	05YQX9_NOCFA	05YQX9 nocardia fa
40	415.5	11.8	258	2	050111_MYCUB	050111 mycobacteri
41	415	11.8	510	2	04NB36_9MICC	04NB36 arthrobacte
42	414.5	11.8	248	2	041WA3_AZOV1	041WA3 azotobacter
43	413.5	11.6	235	2	0989T9_RHIZO	0989T9 rhizobium l
44	406	11.6	236	2	092NP2_RHIME	092NP2 rhizobium m
45	404	11.5	79	2	077242_MYCBO	077242 mycobacteri

## ALIGNMENTS

RESULT 1  
053475 MYCTU  
ID 053475 MYCTU PRELIMINARY; PRT; 681 AA.  
AC 053475\_07D7L3;  
DT 01-JUN-1998 (Tremblrel. 06, Created)  
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
DT 13-SEP-2005 (Tremblrel. 31, Last annotation update)  
DE Hypothetical protein.  
GN OrderedLocustNames=MT2089, Rv2030c;  
OC Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;  
OC Mycobacterium tuberculosis complex.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=H37RV;  
RX MEDLINE=96295987; PubMed=9634230; DOI=10.1038/31159;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C.M.,  
RA Harris D.E., Gordon S.V., Eigmeier K., Gae S., Barry C.E. III,  
RA Tekaita F., Badcock K., Baaham D., Brown D., Chillingworth T.,  
RA Connor R., Davies R.M., Devlin K., Felwell T., Gentles S., Hamlin N.,  
RA Holtroyd S., Hornsby T., Jagers K., Krogh A., McLean J., Moule S.,  
RA Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,  
RA Rogers J., Rutter S., Seeger K., Skelton S., Squares R.,  
RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Disciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544(1998).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=CDC 1551 / Oshkosh;  
RX MEDLINE=2206494; PubMed=12218036;  
RX DOI=10.1128/JB.184.19.5479-5490.2002;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J.D., Deboy R.T., Dodson R.J., Gaitan M.L., Haft D.H.,  
RA Hickey B.K., Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D.,  
RA Salzberg S.L., Delcher A.L., Uterback T.R., Weidman J.F., Khouri H.M.,  
RA Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,  
RA Fraser C.M.;  
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains."  
RL J. Bacteriol. 184:5479-5490(2002).  
CC -1- SUBUNIT: Homodimer (By similarity).  
CC -1- SUBUNIT: Homodimer (By similarity).  
CC -1- SUBUNIT: Homodimer (By similarity).  
EMBL: BX842578; CAI17244.1; -; Genomic DNA.  
EMBL: AB000516; AA46368.1; -; Genomic DNA.  
PIR: E70942; E70942.  
DR TIGR; MT2089; -;  
DR Tuberculist; Rv2030c; -;  
DR GO: GO:0016789; F:carboxylic ester hydrolase activity; IEA.  
DR GO: GO:0016757; F:transferase activity, transferring glycoyl. . .; IEA.  
DR GO: GO:0005116; P:nucleoside metabolism; IEA.  
DR GO: GO:0006166; P:purine ribonucleoside salvage; IEA.

DR GO: 0046677; P: response to antibiotic; IEA.  
 DR InterPro; IPR007815; Erythro. esterase.  
 DR InterPro; IPR000836; P: transferase.  
 DR Pfam; PF05139; Erythro. esterase; 1.  
 DR Pfam; PF00156; Erythro. esterase; 1.  
 KM Complete proteome; Glycosyltransferase; Hypothetical protein;  
 KM Transferase.  
 SQ SEQUENCE 681 AA; 74930 MW; 82DAC6ACC550947 CRC64;  
 Query Match 99.9%; Score 3505; DB 2; Length 681;  
 Best Local Similarity 99.9%; Pred. No. 3e-206;  
 Matches 680; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VLMTAADVTTRSPRRVFRDRREAGRVLAELAAAYRDPDVIYVGLARGGLPVAWEVAAA 60  
 DB 1 MMTAAADVTTRSPRRVFRDRREAGRVLAELAAAYRDPDVIYVGLARGGLPVAWEVAAA 60  
 QY 61 LHAFLDAFVVRKLGAPGHDEFPAVALASGGRVVNDVVRGLRTTPQQLRDIAEREGREL 120  
 DB 61 LHAFLDAFVVRKLGAPGHDEFPAVALASGGRVVNDVVRGLRTTPQQLRDIAEREGREL 120  
 QY 121 LRRESAARGRPPTDITGKTVIVDDGLATGASMFPAVQALRDAPQIVIAVPAPEST 180  
 DB 121 LRRESAARGRPPTDITGKTVIVDDGLATGASMFPAVQALRDAPQIVIAVPAPEST 180  
 QY 121 LRRESAARGRPPTDITGKTVIVDDGLATGASMFPAVQALRDAPQIVIAVPAPEST 180  
 DB 121 LRRESAARGRPPTDITGKTVIVDDGLATGASMFPAVQALRDAPQIVIAVPAPEST 180  
 QY 181 CREPAGLVDDVCAITMPTPLAVGESFMDPROVTDDEVRLLATPTAGSLRRPAASTAA 240  
 DB 181 CREPAGLVDDVCAITMPTPLAVGESFMDPROVTDDEVRLLATPTAGSLRRPAASTAA 240  
 QY 241 DVLRVAIDAPGVPFHEVLAELVGDARIVLIGSSHGTHFVQARAAMTQMLIEBKFG 300  
 DB 241 DVLRVAIDAPGVPFHEVLAELVGDARIVLIGSSHGTHFVQARAAMTQMLIEBKFG 300  
 QY 301 AVAAEADMPDARYRNRVYRGLEEDTNADALSGFERFPAMMMRNTVVRDFEVLRTNR 360  
 DB 301 AVAAEADMPDARYRNRVYRGLEEDTNADALSGFERFPAMMMRNTVVRDFEVLRTNR 360  
 QY 361 YESGALROAGFYGLDLYSLHRSIQEVIISYLDKVPRAAARARAYACFDHACADGQAYG 420  
 DB 361 YESGALROAGFYGLDLYSLHRSIQEVIISYLDKVPRAAARARAYACFDHACADGQAYG 420  
 QY 421 PAAAFGAGPSCEREAVALVQVGNALAYARQDGLADELFYAQONAAQVTRDAEVYRA 480  
 DB 421 PAAAFGAGPSCEREAVALVQVGNALAYARQDGLADELFYAQONAAQVTRDAEVYRA 480  
 QY 481 MFSGRVTSNLRDQMAQTGSLTLTHDRHLDAPPAIVVAHNSHVGARATEVWADGQ 540  
 DB 481 MFSGRVTSNLRDQMAQTGSLTLTHDRHLDAPPAIVVAHNSHVGARATEVWADGQ 540  
 QY 541 LTTGQIVRERYGDSRSIGFSTYTGVTAASEWGGIAQKAVRPALHGSVEELFHQTADS 600  
 DB 541 LTTGQIVRERYGDSRSIGFSTYTGVTAASEWGGIAQKAVRPALHGSVEELFHQTADS 600  
 QY 601 FLVSARLSRDAEALDVVRGLRAIGVYVLPATERQSHYLRPADQPDAMIHIQDTALE 660  
 DB 601 FLVSARLSRDAEALDVVRGLRAIGVYVLPATERQSHYLRPADQPDAMIHIQDTALE 660  
 QY 661 PLEVTSRWINGENETYPYTG 681  
 DB 661 PLEVTSRWINGENETYPYTG 681  
 RESULT 2  
 ID 07T241 MYCBO PRELIMINARY; PRT; 606 AA.  
 AC 07T241;  
 DT 01-OCT-2003 (Tremblrel. 25, Created)  
 DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)  
 DE 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
 DE Hypothetical protein Mb2056c.  
 GN OrderedlocusNames=Mb2056c;  
 OS Mycobacterium Bovis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;  
 OC Mycobacterium tuberculosis complex.  
 OX NCBI\_TaxID=1765;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=AF2122/97;  
 RX MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;  
 RA Garnier T., Eigelmeier K., Camus J.-C., Medina N., Mansoor H.,  
 RA Pryor M., Dutfoy S., Grondin S., Lacroix C., Monsempé C., Simon S.,  
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
 RT "The complete genome sequence of Mycobacterium bovis."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 DR EMBL; BX248341; CAD96909.1; -; Genomic DNA.  
 DR GO: GO:0016789; F:carboxylic ester hydrolase activity; IEA.  
 DR GO: GO:0016757; F:transferase activity; transferring glycosyl. . .; IEA.  
 DR GO: GO:0009116; P:nucleoside metabolism; IEA.  
 DR GO: GO:0006166; P:purine ribonucleoside salvage; IEA.  
 DR GO: GO:0046677; P:response to antibiotic; IEA.  
 DR InterPro; IPR007815; Erythro. esterase.  
 DR Pfam; PF05139; Erythro. esterase; 1.  
 DR Pfam; PF00156; Erythro. esterase; 1.  
 KM Complete proteome; Glycosyltransferase; Hypothetical protein;  
 KM Transferase.  
 SQ SEQUENCE 606 AA; 66540 MW; C9E2278908BF4747 CRC64;  
 Query Match 86.1%; Score 3022; DB 2; Length 606;  
 Best Local Similarity 99.8%; Pred. No. 1e-176;  
 Matches 586; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VLMTAADVTTRSPRRVFRDRREAGRVLAELAAAYRDPDVIYVGLARGGLPVAWEVAAA 60  
 DB 1 MMTAAADVTTRSPRRVFRDRREAGRVLAELAAAYRDPDVIYVGLARGGLPVAWEVAAA 60  
 QY 61 LHAFLDAFVVRKLGAPGHDEFPAVALASGGRVVNDVVRGLRTTPQQLRDIAEREGREL 120  
 DB 61 LHAFLDAFVVRKLGAPGHDEFPAVALASGGRVVNDVVRGLRTTPQQLRDIAEREGREL 120  
 QY 121 LRRESAARGRPPTDITGKTVIVDDGLATGASMFPAVQALRDAPQIVIAVPAPEST 180  
 DB 121 LRRESAARGRPPTDITGKTVIVDDGLATGASMFPAVQALRDAPQIVIAVPAPEST 180  
 QY 181 CREPAGLVDDVCAITMPTPLAVGESFMDPROVTDDEVRLLATPTAGSLRRPAASTAA 240  
 DB 181 CREPAGLVDDVCAITMPTPLAVGESFMDPROVTDDEVRLLATPTAGSLRRPAASTAA 240  
 QY 241 DVLRVAIDAPGVPFHEVLAELVGDARIVLIGSSHGTHFVQARAAMTQMLIEBKFG 300  
 DB 241 DVLRVAIDAPGVPFHEVLAELVGDARIVLIGSSHGTHFVQARAAMTQMLIEBKFG 300  
 QY 301 AVAAEADMPDARYRNRVYRGLEEDTNADALSGFERFPAMMMRNTVVRDFEVLRTNR 360  
 DB 301 AVAAEADMPDARYRNRVYRGLEEDTNADALSGFERFPAMMMRNTVVRDFEVLRTNR 360  
 QY 361 YESGALROAGFYGLDLYSLHRSIQEVIISYLDKVPRAAARARAYACFDHACADGQAYG 420  
 DB 361 YESGALROAGFYGLDLYSLHRSIQEVIISYLDKVPRAAARARAYACFDHACADGQAYG 420  
 QY 421 PAAAFGAGPSCEREAVALVQVGNALAYARQDGLADELFYAQONAAQVTRDAEVYRA 480  
 DB 421 PAAAFGAGPSCEREAVALVQVGNALAYARQDGLADELFYAQONAAQVTRDAEVYRA 480  
 QY 481 MFSGRVTSNLRDQMAQTGSLTLTHDRHLDAPPAIVVAHNSHVGARATEVWADGQ 540  
 DB 481 MFSGRVTSNLRDQMAQTGSLTLTHDRHLDAPPAIVVAHNSHVGARATEVWADGQ 540  
 QY 541 LTTGQIVRERYGDSRSIGFSTYTGVTAASEWGGIAQKAVRPALH 587  
 DB 541 LTTGQIVRERYGDSRSIGFSTYTGVTAASEWGGIAQKAVRPALH 587



```
Query Match      33.1%; Score 1160.5; DB 2; Length 495;
Best Local Similarity 50.7%; Pred. No. 8.3e-63;
Matches 243; Conservative 64; Mismatches 135; Indels 37; Gaps 10;

QY 229 PSLERRPAASTA-----ADVLRRVAI---DAPGVPTHEVLAVLDARIVL 271
DB 28 PGRRTPRSSRSRSPIGMKEEKTMBERDITTRAVASAIIEPLKDGTRMDLALLIGDARIVL 87
QY 272 IGESSHGTHEFYQARAAMTWLIEBKGAVAAEADWDAYVNRVYRGAGEEDTNADEAL 331
DB 88 LGEASHGTHEFYAARAGITQRLITEKGDVAVEADWDPSLRASRYAGSGDDADANAL 147
QY 332 SGFERPPAMMRRNTVVRDFTVWLTNRQRYSGALRQAGFYGLDLYSLHNSIQEYISYLD 391
DB 148 GGFPRFPMWRNTEVAFIEMLRQVNS--SRGADERVGFGLDLYSLRASMDDVVRYLE 205
QY 392 KVDPRAARARARACPDHACADGQAYGFAAAGAGSGCREAVEOLVDYQRNALAYAR 451
DB 206 TVDEBAKRARARACPDH-IAEDPQRYGATTTGTVHEHCEREVLRLQALAEITGDPERYLH 264
QY 452 QDGLAABDELFYAQQNQTVAADAEVYRAMPFSGRVTSMNRDQMAQTLGSLTLHLDRHL 511
DB 265 HDGIAAADELFYAQQNARVANAARAEYRSMFAGHESWNRDASHMAETLEMLGHLDS- 323
QY 512 DAPPARIVVMAHNSHVGDAATEVWADGQLTGQIVNERVDE-SRSIGESTYTYTAA 570
DB 324 RGRPAKIAVMAHNSHVGDAARTEWGEAGEHNLGQLVROFGAEHARLGFTHAGTYAA 383
QY 571 SEMGIIQKRVKVRPALGSGVEELFHOQ-ADSFVSA-----RLSDAEAPLDVNLGRAIG 625
DB 384 SDMSWSPFLKVRKVRSHSPESFERVHDIGVERFLASDRVPLRREB-----RLERRAIG 436
QY 626 VVYLPATEROSHYLVHPADQFDMIHIDQTRALEPLVTSRMTAGENP---ETYPYGL 681
DB 437 VIYLPGERVSHYFYADIAKQFDAVIHDDRRLALPLDPSITMWSGSSPEQMERTYPSGM 495

RESULT 6
Q4HTU3_9DEIO PRELIMINARY; PRT; 448 AA.
ID 04HTU3_9DEIO PRELIMINARY; PRT; 448 AA.
AC 04HTU3_
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Erythromycin esterase.
GN ORFNames=DgEDRAFT_0443;
OS Deinococcus geothermalis DSM 11300.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=319795;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=DSM 11300;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Deinococcus geothermalis
RT DSM 11300."
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=DSM 11300;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Deinococcus geothermalis
RT DSM 11300."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; RAHE01000008; BA82343.1; -, Genomic DNA.
SQ SEQUENCE 448 AA; 50995 MW; 6E3EEFE136D64750 CRC64;
```

```
Query Match      31.9%; Score 1120.5; DB 2; Length 448;
Best Local Similarity 53.1%; Pred. No. 2.1e-60;
Matches 234; Conservative 62; Mismatches 134; Indels 11; Gaps 8;

QY 243 LRRVADPAGVPTHEVLAVLDARIVLIGESSHGHTEFYQARAAMTWLIEBKGAV 302
DB 15 LRDVAPRLTGAADYDLDLDCIGNARFVLIGESSHGHTEFYRRARLTMRLIEKEFTAV 74
QY 303 AAEADWDAYVNRVYRGAGEEDTNADEALSGFERPPAMMRRNTVVRDFTVWLTNRQRYE 362
DB 75 AVEADWDAYVNRVYRGAGEEDTNADEALSGFERPPAMMRRNTVVRDFTVWLTNRQRYE 133
QY 363 SGALRQAGFYGLDLYSLHNSIQEYISYLDKVDPRRAARARACPDHACADGQAYGFA 422
DB 134 PGA--PVGFGLDLYSLHNSMAVVELETVDDEAARARQVGCDFQ--FGENPOAGYA 190
QY 423 AARGAGSGCREAVEOLVDYQRNALAYARDDGLAEDELFYAQQNQTVAADAEVYRAMPF 482
DB 191 TESGRREPCDDAAVQQLLELQRRBA--QESGFLAEDEDFYQARNAKLAKAETTYRAMF 248
QY 483 SGRVTSMNLRDQMAQTLGSLTLHLDRHLDAPPARIVVMAHNSHVGDAATEV-WADGQL 541
DB 249 RGRDESWSLDAMWAELELVEHGER--QGRPOKIVVMAHNSHVGDAARASEMGWLRGEL 306
QY 542 TLGQIVNERVDE-SRSIGESTYTYTAASEKGIQKRVKVRPALGSGVEELFHOQAD-S 600
DB 307 NVQQLAREPGRFTFYVQSTHNGTYAADDWDEPVARVARPALPESVVDLLHEVGEAA 366
QY 601 FVLSARLSRAEAPLDVNRGIRAGVYVLPATEROSHYLVHPADQFDMIHIDQTRALE 660
DB 367 YMLDKRENPATGLRARELRQRFQIYIRPETERMSHYITRLSDMYDALLPFDETSAYV 426
QY 661 PLEVTS-RWJAGENPETYPYTG 680
DB 427 PLDATAGTEPEGEVPDTPFTG 447

RESULT 7
Q089L97_BRAJA PRELIMINARY; PRT; 679 AA.
ID 089L97_BRAJA PRELIMINARY; PRT; 679 AA.
AC 089L97;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Protein-L-isoaspartate O-methyltransferase.
GN OrderedLocusNames=b114651;
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=USDA 110;
RG MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idegawa K., Iritschi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsurunoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
CC -1- FUNCTION: Catalyzes the methyl esterification of L-isoaspartyl
CC residues in peptides and proteins that result from spontaneous
CC decomposition of normal L-aspartyl and L-asparaginyl residues. It
CC plays a role in the repair and/or degradation of damaged proteins
CC (by similarity).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + protein L-
CC isoaspartate = S-adenosyl-L-homocysteine + protein L-isoaspartate
CC alpha-methyl ester.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
DR EMBL; BA000040; BAC49916.1; -, Genomic DNA.
DR HSP; Q81ZK3; 1UG1.
```



DR GO:0016789; F:carboxylic ester hydrolase activity; IEA.  
 DR GO:0004719; F:protein-L-isoaspartate (D-aspartate) O-meth. .; IEA.  
 DR GO:0006464; P:protein modification; IEA.  
 DR GO:0046677; P:response to antibiotic; IEA.  
 DR InterPro: IPR007815; Erythro\_esteras.  
 DR InterPro: IPR00682; PCMT.  
 DR InterPro: IPR00051; SAM\_bd.  
 DR Pfam: PF05139; Erythro\_esteras; 1.  
 DR Pfam: PF01135; PCMT; 1.  
 DR TIGRfams: TIGR0080; pmtc; 1.  
 DR PROSITE: PS01279; PCMT; 1.  
 KM COMPLETE proteome; Methyltransferase; Transferase.  
 SQ SEQUENCE 679 AA; 75244 MW; AFE5B82707429338 CRC64;

Query Match 27.3%; Score 956.5; DB 2; Length 679;  
 Best Local Similarity 37.3%; Pred. No. 3.8e-50;  
 Matches 259; Conservative 72; Mismatches 207; Indels 157; Gaps 22;

QY 24 AGRVLAELAAVRDQPDVIVIGLARGGLPVAMEVAAL-----H 62  
 DB 105 AAATVSEIAA---DYVTVRLG-----PLAEBAATLTGLGYDVAHVLAGDGTGMDEH 155  
 QY 63 APLAFVVRKVGAPHDPEFVAGALASGGRVV-----NDVVRGLRITPQQLR--DIA 113  
 DB 156 APYDAIVV-AAGGPVPALKEQLKIGRLVIVPGADQRTQELVAVVRLSKDEFSESDIA 214  
 QY 114 EREGSELRLRE--SAVRGERPPTDTTGKTVIVDDGLATGASMPFAVVALDAQPAQVI 171  
 DB 215 DVRFVPLIGEGGMAAKGTAP-----ARALK----- 242  
 QY 172 AVPAPESTCEBFAGLVDDVCATMPTEFLAVGSEFMDPROVTDEE--VRRLATPTAGP 229  
 DB 243 --PVAP-----DEETLVRIADAASFP 263  
 QY 230 SLRRPASTADVLRRVAIDARGVPTHEVLAEVLGDAHYLLIGSSHGTEFGYQARAM 289  
 DB 264 SIE-----AADL-----SPMERIGSARVVLGATGTSSEFYMRRRI 302  
 QY 290 TQWLIIEKFGVAEAADMDPAYRVNRYRGISGDEPTNDAELSGFERPAMMMRTVVRD 349  
 DB 303 TRDLIVKGFPRFVAIEADMPDARVDHYVRF--QYPSSEWTAFARFPTMMWMTTEVRD 359  
 QY 350 FVEWIKTRNQYESGALRQAGFYGLDLYSLHSIQEIVSYLDKVDPRRAADPARARYCED 409  
 DB 360 FVSWLRKINGTVEN--TRVAFHGIDLXSLYDSIRSVINLYDEVDPASARVARREYGLT 417  
 QY 410 HACADDCGAYGPAAGFGSPCEBAVAGQVLDVQNMALAYARODGLAEDLFTYQOQAQ 469  
 DB 418 -PWQRDPATYGHAAITGSPPTCESDVAAHALTDLAKRRAVVAEHQD---ERFLDEQNAK 472  
 QY 470 TVRDAEVYRRAMSGRVTSMILRDOHMAQTIGSLTLTHLRHLADAPPAIVVAHNSHYGD 529  
 DB 473 LVANAERYRYRIMYIGSRASMNLRDSHMFDTLKNLLA-----FHGDSQAVVAHNSHYGN 527  
 QY 530 ARATEWADGQLITGQIVRERYGDESRSIGFSTYGTIVTAASEWGIQAKRAVPAALGS 589  
 DB 528 AATTEMVARGEHNIGQHCRKAFGAQAYLVGFTHSGTVAASDMGSPMEVTVTRSLNS 587  
 QY 590 VEEFLHQTA--DSFLVASLSDAENP--LDVVRLEGRAIGVYVLPTEQSHYLAHVRPDQ 646  
 DB 588 YERLCHAATGLPRFMGLRGPGDLCPBGIGRELERAGVYIRPETELASHYFQASLPQQ 647  
 QY 647 FDAMIHIDQTRALBPLEVTSRWIAGENPEYPTGL 681  
 DB 648 FDEYIWFDDTRAVTPLETAE--IAG-LPDMTIPFGV 679

RESULT 8  
 O6M168 RHISN  
 ID O6M168 RHISN PRELIMINARY; PRT; 653 AA.  
 AC O6M168  
 DT 05-JUL-2004 (Tremblrel. 27, Created)  
 DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)

DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
 DB Protein-L-isoaspartate O-methyltransferase (EC 2.1.1.77).  
 GN ORNames=NGR00375;  
 OS Rhizobium sp. (strain NGR234).  
 OG Plasmid megaplasmid 2.  
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 CC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=394;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=NGR234;  
 RA Broughton W.J., Perret X., Staehelin C., Schmitz R.A., Raesch C.,  
 RA Liesegang H., Gottschalk G., Streit W.R.;  
 RT "Comparative DNA analysis of two large contigs of the Rhizobium sp.  
 NGR234 megaplasmid 2.";  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 CC -I FUNCTION: Catalyzes the methyl esterification of L-isoaspartyl  
 residues in peptides and proteins that result from spontaneous  
 decomposition of normal L-asparaginyl and L-asparaglyl residues. It  
 plays a role in the repair and/or degradation of damaged proteins  
 (by similarity).  
 CC -I CATALYTIC ACTIVITY: S-adenosyl-L-methionine + protein L-  
 isoaspartate = S-adenosyl-L-homocysteine + protein L-isoaspartate  
 CC alpha-methyl ester.  
 CC -I SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC EMBL: AY116747; AA087500.1; -; Genomic DNA.  
 DR GO:0016789; F:carboxylic ester hydrolase activity; IEA.  
 DR GO:0004719; F:protein-L-isoaspartate (D-aspartate) O-meth. .; IEA.  
 DR GO:00000179; F:RNA (adenine-N6,N6)-dimethyltransferase a. .; IEA.  
 DR GO:0008649; F:RNA methyltransferase activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006464; P:protein modification; IEA.  
 DR GO:0046677; P:response to antibiotic; IEA.  
 DR GO:0000154; P:RNA modification; IEA.  
 DR InterPro: IPR007815; Erythro\_esteras.  
 DR InterPro: IPR000682; PCMT.  
 DR InterPro: IPR001737; RNA\_meth\_trans.  
 DR InterPro: IPR00051; SAM\_bd.  
 DR Pfam: PF05139; Erythro\_esteras; 1.  
 DR Pfam: PF01135; PCMT; 1.  
 DR SMART: SMO0650; RADC; 1.  
 DR TIGRfams: TIGR0080; pmtc; 1.  
 DR PROSITE: PS01279; PCMT; 1.  
 KM Methyltransferase; Plasmid; Transferase.  
 SQ SEQUENCE 653 AA; 72564 MW; E2BFF829C764E511 CRC64;

Query Match 26.2%; Score 919; DB 2; Length 653;  
 Best Local Similarity 37.4%; Pred. No. 7.3e-48;  
 Matches 257; Conservative 66; Mismatches 228; Indels 136; Gaps 23;

QY 6 AADYTRSPRRVF---BDRREAGRVLAELAAVRDQPDVIVIGLARGGLPVAMEVAALH 62  
 DB 92 AAATVSEIAA---DYVTVRLG-----PLAEBAATLTGLGYDVAHVLAGDGTGMDEH 142  
 QY 63 APLDAFVVRKVGAPHDPEFVAGALASGGRVVVNDVVRGLRITPQQLRDAEREGRELL- 121  
 DB 143 APFDALIV-----AAGGPV-----PRALKEOLEIGGRLLIP 174  
 QY 122 --RESAYRGERPPTDTTGKTVIVDDGLATGASMPFAVVALDAQPAQVIIVAPAPES 179  
 DB 175 VGRBOQAQRLLR---IT-----RTAANRY-----EEEDLGGVLFPP----- 207  
 QY 180 TCRFAGLVDDVCATMPTEFLAVGSEFMDPROVTDEEVRLLATPTAGSLRRPASTA 239  
 DB 208 -----LIGEGMSSESPDHARRV-----PDL-----VAEA 235  
 QY 240 ADVLRVAIDAPGVPTHEVLAEVLGDAHYLLIGSSHGTEFGYQARAMTQWLIIEKGF 299  
 DB 236 ADDLPLVT--DPGADPFERFAR---HRIVLGEATHGTSSEFYQARAIVTORLIERHGF 289  
 QY 300 GAVAAEADMDPAYRVNRYRGISGDEPTNDAELG---FERPAMMMRTVVRDQVLEKLT 356  
 DB 290 AIVAVEADMDPAYRVNRYRGISGDEPTNDAELG---FERPAMMMRTVVRDQVLEKLT 342



KM	Complete proteome; Hypothetical protein; Methyltransferase; Plasmid; Transferase.	
5Q	SEQUENCE 658 AA; 73849 MW; CCC50A45F3331CD2 CRC64;	
Query Match	25.5%; Score 894.5; DB 2; Length 658;	
Beet Local Similarity	36.7%; Pred. No. 2.3e-46;	
Matches	255; Conservative 81; Mismatches 269; Indels 89; Gaps 23;	
QY	36 RDGDPVIVTGLARGGLPFAWEEVAALALAPDAFVVRKLGAPGHDFAV--GALASGRVV 93	
DB	6 RAREMTTSHISRGRIDRHRHVAEMGVPRAPF---DPGEFAYEDSAUSISHGQT 60	
QY	94 VNDVVGGLRTTPOLR-----DIAREG--RELLRRESAY-----RGERPP 133	
DB	61 ISGPYIYALMIERAEVQPGDTVEIGTSGVAAAVLSRIAAVYVTRHAGLAEVARRF 120	
QY	134 TDIT-GKTIVVDGGLA--TGASMPAAV-----QALRDAQPAQIYIYANPAEST 180	
DB	121 AELRYGNDIVRGVGGTGWPEAGPFDAILVAAGSEPIPHALKEDLDGHLVIVPGPEE 180	
QY	181 CREFAGLVDVDVCAETMPFLA-----VGSEFMDFROYTDECVRLRLATPTGAPSLRR 233	
DB	181 QRLMK--VTRNATTFEEQDLGSRFPPLVGEYK-----HEERAGSRAPRPAPFLPE 232	

Db	223	LVAE--AAEPPLD-----DDPARGLLPDRPAGRRIVLIDEGASHGSEFYRARAALTRRL	265
Qy	294	IEEYGFGAANAADMPDAIKYNRIVYRGLGEBTPNADDELSCFERPPAMRNTVYRDDEVEN	353
Db	286	IEEHGFALVAEADMPAAAVDRVYRRHQIRLD--APQRFPTWMNRREVMDFEEN	342
Qy	354	LRTNQRYESGALRQAGFYGLDLYSLRHSIOEVI SYDKYDPAPRAAARARAYACFDHACA	413
Db	343	MREHNGR--RSLPDRAGFYGLDLYNMGSGISAVLRIYDEIDPEAAAVARRRYGLT--PMQ	399
Qy	414	DDGQAYGPAAFGAGPSCEREAEQYLDVQDNALAYARQDGLAEDLFFAQQNAQTVRD	473
Db	400	NEPSTYGAAMTAGFRKCEBAVYVQRCHELLEKOLEAGRDGSG--DELDLDAVQNAFLVAS	455
Qy	474	AEVYVYAMPFSGRIVYKMLRDQMAQOTGSLTTHLDRLDAPPAIRIYVYMANSHVGDARAT	533
Db	456	AERYIRIMYAYAGADSKMNRDTHMFTETLHLL-----NARGARSAVYVMAHNSHIGDARHT	510
Qy	534	EV-WADQQLTLGQIVRERYGDESHSIGSFYTYGTVTYVTAASEWGGIAQKAVRPAALHGSVEE	592
Db	511	DMGARAREELNIQGLCRERFGDQALLIGFTGHGVAASAQMNQDMVEKYPRIPLSEGSYER	570
Qy	593	LFHQT-ADSPFVS----ALSRDRAEALDVYRLGRALGVVYVLTATEQSHYLVLRPAPDOF	647
Db	571	VMDHSGVGCEFLIDPARRARLRDGLKLPL---LERRFGVITVRPDTERRSHYALASLPRQF	626
Qy	648	DAMTHIDOTRALPLEVTSRWIAGENBETPYTGL 691	
Db	627	DGFVMPDQTPTRVEPL--GAQHIAKGVDPDTPPEGL 658	
RESULT 10			
Q6MCK4 PARUM			
ID	Q6MCK4	PARUM PRELIMINARY;	PRT; 456 AA.
AC	Q6MCK4;		
DT	05-JUN-2004	(TREMBLRel. 27, Created)	
DT	05-JUL-2004	(TREMBLrel. 27, Last sequence update)	
DT	05-JUL-2004	(TREMBLrel. 27, Last annotation update)	
DE	Hypoetical protein.		
CN	OrderedLocustNames=pc0851;		
OS	Parachlamydia sp. (strain W6235) (subsp. Acanthamoeba sp.).		
OC	Bacteria; Chlamydiae; Chlamydiales; Parachlamydiaceae; Parachlamydia.		
OX	NCBI_TaxID=264201;		
RM	[1]		
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].		
RA	PubMed=15073324;		
RA	Horn M., Collingro A., Schmitz-Baser S., Beier C.L., Purkhold U.,		





```
RESULT 14
Q4WJ15 ASPFU PRELIMINARY; PRT; 463 AA.
ID Q4WJ15
AC Q4WJ15
DT 13-SEP-2005 (TREMBLrel. 31, Created)
DT 13-SEP-2005 (TREMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TREMBLrel. 31, Last annotation update)
DE Erythromycin esterase family protein.
GN ORFNames=Afu105850;
OS Aspergillus fumigatus Af293.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxId=330879;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Af293;
RA Niernan W., Pain A., Anderson M.J., Wortman J., Kim H.Stanley.,
RA Atroya J., Britman M., Abe K., Archer D.B., Berneto C., Bennett J.,
RA Boyer P., Chen D., Collins M., Coulson R., Davies R., Dyer P.S.,
RA Farman M., Pedorova N., Pedorova N., Feldlyum T.V., Fischer R.,
RA Fokker N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,
RA Goldman G.H., Gomi K., Griffith-Jones S., Gilliam R., Haas B.,
RA Haas H., Harris D., Horiuchi H., Huang J., Humphrey S., Jimenez J.,
RA Keller N., Knout H., Kitamoto K., Kobayashi T., Kulkarni R.,
RA Kumagai T., Lafton A., Latge J.-P., Li W., Lord A., Lu C.,
RA Majoros W.H., May G.S., Miller B.L., Mohamoud Y., Molina M., Monod M.,
RA Moulay I., Mulligan S., Murphy L., O'Neill S., Paulsen I.,
RA Penaya M.A., Perteu M., Price C., Pritchard B.L., Quail M.A.,
RA Rabinowitsch E., Rawlins N., Rajandream M.-A., Reichard U.,
RA Renauld C.M., Rutter S., Salzberg S.L., Sanchez M.,
RA Sanchez-Ferrero J.C., Saunders D., Seeger K., Squares R., Squares S.,
RA Takenuchi M., Tekala F., Turner G., Vazquez de Aldana C.R., Weidman J.,
RA White O., Woodward J., Yu J.-H., Fraser C., Galagan J.E., Asai K.,
RA Machida M., Hall N., Barrett B., Denning D.W.;
RT "Genomic sequence of the pathogenic and allergenic filamentous fungus
RT Aspergillus fumigatus";
RL Submitted (May-2005) to the EMBL/Genbank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AAFH0100007; EAL88297.1; -!- Genomic DNA.
DR EMBL; AAFH0100007; EAL88297.1; -!- Genomic DNA.
SQ SEQUENCE 463 AA; 53158 MW; CB4E1B508472D202 CRC64;

Query March 23.1%; Score 811.5; DB 2; Length 463;
Best Local Similarity 42.0%; Pred. No. 1.8e-41;
Matches 198; Conservative 66; Mismatches 172; Indels 35; Gaps 16;

QY 230 SLRRPASTADVADRVVADAPG-GVPTHEVLAELVGDARIIVLIGSSSHGTHFYQARA 288
DB 9 SSRPFWMAQLQOLQTEAAQPIPTISDPSFGSHFNFNGYKVVLLGDSHGSEFYAARAE 68
QY 289 MTQWLIBEKGFAVAEAADWDPAVVRNRYV-----GLGEDTNADALSGFERPAMM 341
DB 69 ITRKLIEHQGYTVALEADWDPAEDRIDRYVRQPGPKAGIS---GKADLEPPFKFPYWM 125
QY 342 WRNTVAVDFTVWLTTRNRYESSGALRQGFGLDLSLRHSIQEVIYSLDKVDPRAARA 401
DB 126 WRNEMODLVEMERNRAKLPNE--KAGVYGLDLSGASIRAVITDLVDVDPAGKEA 183
QY 402 RARAFCEFDHACADQOAGYGAFAAGAGPSGCEAEVQDVQORNALAARODGLLAEDEL 461
DB 184 RRRIGCL-QPWPVDEPSAYGLASLRGM--EDCESGVLCMKRLDLLKRLQYAGND--VRDGE 239
QY 462 FYA-QQNAQTVRDAEVYVRAMFSGRVTSWNLRDQMAQTGLSLTLDRHLADAPRATV 520
DB 240 FHSSEQNAFVVRDAERYRYKAYYSASASWTLRDTHMFDTLRLFRH-----KPAQAKIV 294
QY 521 WAHSHVGDAAATEVMA-DGQITLIGQIVRERYGDSRSI-GFSYTYGTVTAASWGGIAQ 578
DB 295 WAHSHGCDAAATYMGTRNREVNIGQLIRENFGRGENAVALGCGTHGTVAHAHEMDDMQ 354
```

```
QY 579 RKAVRPLHSGVSELEPHQTA-DSFLVSARLS-----RDEAPLDVRLGRAIGVYLP 631
DB 355 VMKRPSPRSDSWETIADTGPISFVILDRKHLDPALRTMA-AENSRERFPGIVYRPD 413
QY 632 TERQSHYLVHVPADQFDMHIIDOTRALPELE-VTSRWIAGENPEYPTGL 681
DB 414 TERISHYSQAYLNQFPAIYMPDTEAVKPELEKVPKPTPLGDP-ETTFPFGV 463

RESULT 15
Q41713 GIBZE
ID Q41713 GIBZE PRELIMINARY; PRT; 455 AA.
AC Q41713
DT 13-SEP-2005 (TREMBLrel. 31, Created)
DT 13-SEP-2005 (TREMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TREMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=FG06995.1;
OS Gibberella zeae PH-1.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
OX NCBI_TaxId=229533;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PH-1;
RA Birren B., Nussbaum C., Abouelleil A., Allen N., Anderson S.,
RA Archchi H.M., Barua N., Bastien V., Bloom T., Bogulavsky L.,
RA Boukhalter B., Butler J., Calvo S.E., Camarata J., Chang Y.,
RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., Dearellano K.,
RA Diaz U.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,
RA Gadryna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagedorn D., Haos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaffe D., Johnson R., Jones K., Kamal M., Kanat A., Karatas A.,
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Maucelli E., McCarthy M., Meldrum J., Menues L.,
RA Milnova T., Mienga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neill D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Puccell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA Roman J., Schauer S., Schupbach R., Seaman S., Severy P., Smitrov S.,
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.;
RT "Fusarium graminearum genome sequence";
RL Submitted (Feb-2004) to the EMBL/Genbank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AAC01000289; EAA76190.1; -!- Genomic DNA.
DR EMBL; AAC01000289; EAA76190.1; -!- Genomic DNA.
SQ SEQUENCE 455 AA; 51820 MW; C3148AAE899A6D7F CRC64;

Query March 22.8%; Score 800.5; DB 2; Length 455;
Best Local Similarity 42.4%; Pred. No. 8.5e-41;
Matches 185; Conservative 62; Mismatches 156; Indels 33; Gaps 12;

QY 265 GDARIVLIGSSSHGTHFYQARAAMTQWLIBEKGFAVAEAADWDPAVVRNRYV----- 319
DB 34 GDCVKVLLIGASHGTSFYSVRAEITKVTMEHNGFINVAEADWSDAEHVDVRRHRPVP 93
QY 320 --GLGEDT--NADDEALSGERPPAMWRRTTVTRDFTVEMLRTNRYESG--ALRQAGFY 372
DB 94 GQGAVERITQWAEKEKESPLRPTWWRNVEVHDPEWRS-----YNSREVTAAAGFY 149
QY 373 GLDYSLSHRISQEVISYLDVDPRAAARARYA-CFDHACADQOAGYGAFAAGAGPSC 431
DB 150 GLDYSLSGTSKMAVIDVLDVDPKMAKVARGRIYNLMD--WLEDPEHYGLSGLATSKGY 207
QY 432 EREAVEQLVDVQNNALAY-ARODGLLAEDELFYAQQNAQTVRDAEVYVRAMFSGRVTSWN 490
```

```

Db      208 EQDVVAMIGDILRRKRIEYSALDGV---EFHNGEQNARVVKDAEQYKAMYRGODKSWN 264
QY      491 LRQDMHMOITLGSLLTHLDRIHDAPPARIWVWANSNHSVGDARATEV-WADGOLTLGOIVRE 549
Db      265 HRDMHMFETTLKRVLEH-----RGECSKAIWVWANSHTGDADATSMNSHSHLNLIGELCKR 319
QY      550 RYGDERSRSIGFSTYTGIVTAASEWGGIAQRKAVRPALHGSVEELPHQTA-DSPVLSARLS 608
Db      320 AFGDHALSIGTGTWGTVAQAQNWESDWNIIKVQPLGSGYEELMHATGIGNFVLDLRKG 379
QY      609 RDAEAPLDV---PLGRAIGVYVLPATEROSHYLHVPRADQPDAMIHIDQTRALEPLEVT 665
Db      380 KCDEKLREPALNGERLERSIGVIYKPETERKASHYSYAILPDQDGYWDESKAVGTLEIH 439
QY      666 SRWLAGENPETYPGTL 681
Db      440 QPRSPLEVHETWPPGL 455

```

Search completed: March 23, 2006, 05:24:20  
 Job time : 177.174 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 05:24:53 ; Search time 46.4465 Seconds  
(without alignment)  
1212.194 Million cell updates/sec

Title: US-10-617-038-23  
Perfect score: 3508  
Sequence: 1 VLMTAADVTRSPRRVFRD.....LEVTSMWAGENPEPTPTGL 681

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/6.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/H.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/PCUTUS.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/RE.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	547.5	15.6	226	US-09-902-540-10451	Sequence 10451, A
2	417.5	11.9	229	US-09-902-540-15374	Sequence 15374, A
3	415.5	11.8	289	US-08-311-731A-396	Sequence 396, App
4	346.5	9.9	227	US-09-902-540-12846	Sequence 12846, A
5	147.5	4.2	1650	US-09-252-991A-21798	Sequence 21798, A
6	143.5	4.1	2214	US-09-902-540-15988	Sequence 15988, A
7	136	3.9	685	US-09-252-991A-26723	Sequence 26723, A
8	134.5	3.8	3031	US-07-689-008-2	Sequence 2, App11
9	134	3.8	3562	US-09-679-279-14	Sequence 14, App1
10	131	3.7	535	US-09-252-991A-17140	Sequence 17140, A
11	131	3.7	1275	US-08-426-630-49	Sequence 49, App1
12	130	3.7	6396	US-09-410-551B-72	Sequence 72, App1
13	130	3.7	6396	US-09-940-316B-72	Sequence 72, App1
14	128.5	3.7	2841	US-09-477-962-117	Sequence 117, App1
15	128.5	3.7	3413	US-10-042-665A-8	Sequence 8, App11
16	128	3.6	923	US-09-252-991A-22409	Sequence 22409, A
17	127.5	3.6	7257	US-09-325-409-5	Sequence 5, App11
18	127.5	3.6	7257	US-09-568-102-5	Sequence 5, App11
19	127.5	3.6	7257	US-09-567-969-5	Sequence 5, App11
20	127.5	3.6	7257	US-09-568-480-5	Sequence 5, App11
21	127.5	3.6	7257	US-09-568-486-5	Sequence 5, App11
22	127.5	3.6	7257	US-09-568-472-5	Sequence 5, App11
23	127.5	3.6	7257	US-09-567-899-5	Sequence 5, App11
24	127.5	3.6	7257	US-10-014-717-5	Sequence 5, App11
25	127	3.6	1411	US-09-252-991A-23628	Sequence 23628, A
26	126.5	3.6	903	US-09-902-540-11595	Sequence 11595, A
27	126.5	3.6	3739	US-09-320-878-2	Sequence 2, App11

28	126.5	3.6	3739	2	US-09-105-537-33	Sequence 33, App1
29	126.5	3.6	3739	2	US-09-141-908-3	Sequence 3, App11
30	126.5	3.6	3739	2	US-09-657-440-2	Sequence 2, App11
31	126.5	3.6	3739	2	US-09-793-708-2	Sequence 2, App11
32	126.5	3.6	11877	2	US-09-105-537-6	Sequence 6, App11
33	126	3.6	577	2	US-09-902-540-14076	Sequence 14076, A
34	125.5	3.6	1053	2	US-09-902-540-12126	Sequence 12126, A
35	125	3.6	852	2	US-09-902-540-13582	Sequence 13582, A
36	124.5	3.5	2108	2	US-09-252-991A-31502	Sequence 31502, A
37	124	3.5	1875	2	US-10-042-665A-2	Sequence 2, App11
38	124	3.5	1939	2	US-10-152-886-43	Sequence 43, App1
39	123	3.5	879	2	US-09-252-991A-27857	Sequence 27857, A
40	123	3.5	879	2	US-09-252-991A-22604	Sequence 22604, A
41	123	3.5	4472	1	US-08-804-227C-2	Sequence 2, App11
42	122	3.5	729	2	US-09-252-991A-32535	Sequence 32535, A
43	122	3.5	765	2	US-09-252-991A-19778	Sequence 19778, A
44	122	3.5	831	2	US-09-252-991A-30097	Sequence 30097, A
45	121	3.4	1891	1	US-08-804-227C-12	Sequence 12, App1

## ALIGNMENTS

```
RESULT 1
US-09-902-540-10451
; Sequence 10451, Application US/0902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(11549)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 10451
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-10451

Query Match      15.6%; Score 547.5; DB 2; Length 226;
Best Local Similarity 51.8%; Pred. No. 3.9e+46;
Matches 118; Conservative 30; Mismatches 73; Indels 7; Gaps 2;

QY 12 RSPRRVFRDRRAGRVLAELAAVYRQPDVIVIGIARGLPVAVEVAALHAPIDAEVVR 71
DB 2 REPV-FQDRYEGGRVLAADHLRRARPGTIVLALPPGCVPGVAGVARKGVPLDVLVR 59

QY 72 KLAGPGRDEPAVAGALASGGRVVNDVVRGLRTPOQLRDIABERREGLLRBSAYNGER 131
DB KLAGPGRDEPAVAGALASGGRVVNDVVRGLRTPOQLRDIABERREGLLRBSAYNGER 119

QY 132 PPTITKTVIVDDGLATGSMFAVOALRDAQPAIVIVAPPAPESTCRFAGLVDDV 191
DB 120 PPPVTRRTVTLVDDGLATGTYMAAVALRLLEPAIIVAVPAAAESCSGLQVADEV 179

QY 192 VCAMPPTPLAVGESFMDPROVTDDEVRRLATPTAAPSILRRPAASTA 239
DB 180 ICRMPPEFTSVGLMYRDFQATEDDEVRALLAEXT-----REGAASRA 222

RESULT 2
US-09-902-540-15374
; Sequence 15374, Application US/0902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
```

APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15849)B  
CURRENT APPLICATION NUMBER: US/09/902,540  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 15374  
LENGTH: 229  
TYPE: PRT  
ORGANISM: Myxococcus xanthus  
US-09-902-540-15374

Query Match 11.9%; Score 417.5; DB 2; Length 229;  
Best Local Similarity 42.5%; Pred. No. 1,2e-34;  
Matches 96; Conservative 39; Mismatches 86; Indels 5; Gaps 3;

QY 18 FRDRRAGRVLAELL--AAYRDPDVIYVIGLARGGLPVAMEVAAALHAPLDAFVVRKLG 75  
DB 3 FEDRVDAAGRRLAQLLRGGYTG-GLIVLALPRGGVPVAFVATALGAPLDVWVRKGV 61  
QY 76 PGHDEFVAGALASGRVYVNDVVRGLRITPQQLRDIAERGRLLRRESAYRGERPPTD 135  
DB 62 PGYELGIGAVAEQVAFVNNRLMDEVGTEDMQLVRQKTDVKRVAFRQGIAPR 121  
QY 136 ITGTVIVDDGLATGASMPFAVQALRDAPQIVIAVPAPESTCEFAGLVDDVCAT 195  
DB 122 IEGRIITLVDDGATGCGVRAIQLNRHRSIITLAVPAASQTLAELAPLDVVCVL 181  
QY 196 MPTPLAVGESFMDFRQVTDSEVRRLATP--TAPSLRPAASTA 239  
DB 182 STPLVAIGQVYADFQVDPDAVATLLAQRLTLGHDRHAPGDTA 227

RESULT 3  
US-08-311-731A-396  
Sequence 396, Application US/08311731A  
Patent No. 6583266  
GENERAL INFORMATION:  
APPLICANT: SMITH, DOUGLAS  
APPLICANT: MAO, JEN-I  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR  
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 411  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
STREET: 600 ATLANTIC AVENUE  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/311,731A  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: GATES, EDWARD R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: C0044/7125  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/720-3500  
TELEFAX: 617/720-2441  
INFORMATION FOR SEQ ID NO: 396:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 289 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Mycobacterium leprae  
US-08-311-731A-396

Query Match 11.8%; Score 415.5; DB 2; Length 289;  
Best Local Similarity 41.6%; Pred. No. 2,9e-34;  
Matches 97; Conservative 33; Mismatches 90; Indels 13; Gaps 3;

QY 12 RSPRRVFRDRRAGRVLAELLAAAYRDPDVIYVIGLARGGLPVAMEVAAALHAPLDAFVVR 71  
DB 53 RNSMRLEFDDVDVDAGRHLAERLESLSRGK-DWVVLGLPRGGVPVAFVAKALRAPLDVLR 111  
QY 72 KLGAPGHDEFVAGALASGRVYVNDVVRGLRITPQQLRDIAERGRLLRRESAYRGERPPTD 131  
DB 112 KLGAPVPOPELAFGALGEGVTVNDSVVAEADISQEMALITEQRAELRRSERFHRH 171  
QY 132 PPTDITGTVIVDDGLATGASMPFAVQALRDAPQIVIAVPAPESTCEFAGLVDDV 191  
DB 172 DRAPIGRIAVIYDDGATGATKASCQVRAQGVSKVLAVPVGGRDIFARFAGVADV 231  
QY 192 VCATMPTPLAVGESFMDFRQVTDSEVRRL-----ATPAG-PSLR 232  
DB 232 VCLHTPAFFCAVQGYCNFTQTSDAEVIALDRAREGFESATTATLGDPOIR 284

RESULT 4  
US-09-902-540-12846  
Sequence 12846, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15849)B  
CURRENT APPLICATION NUMBER: US/09/902,540  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 12846  
LENGTH: 227  
TYPE: PRT  
ORGANISM: Myxococcus xanthus  
US-09-902-540-12846

Query Match 9.9%; Score 346.5; DB 2; Length 227;  
Best Local Similarity 41.0%; Pred. No. 2,8e-27;  
Matches 84; Conservative 34; Mismatches 84; Indels 3; Gaps 3;

QY 18 FRDRRAGRVLAELLAAAYRDPDVIYVIGLARGGLPVAMEVAAALHAPLDAFVVRKLGABG 77  
DB 2 FRDRADAGRRLAARLLPYRG-GEVRLGLARGGLRAVVAALAEAPLDVWVSRIDVPG 60  
QY 78 HDEFVAGALASGRVYVNDVVRGLRITPQQLRDIAERGRLLRRESAYRGERPPTDIT 137  
DB 61 R-MTVLGAVSEGGIYLDQDALRGGLPEVARSILARASVNDQVQRLRGTSERPA-WG 118  
QY 138 GKTIVIVDDGLATGASMPFAVQALRDAPQIVIAVPAPESTCEFAGLVDDVCATMP 197  
DB 119 GFTVLVDDGLVSGATAMALDYLRRQHPARLVGVPTPHGLARVREDAVHCV 178  
QY 198 TPFLAVGESFMDFRQVTDSEVRRL 222  
DB 179 PAMRDVSEAYVDYRPLVDVLRQL 203

RESULT 5



US-09-252-991A-21798  
; Sequence 21798, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIORITY FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21798  
; LENGTH: 1650  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21798

Query Match 4.2%; Score 147.5; DB 2; Length 1650;  
Best Local Similarity 22.1%; Pred. No. 3.7e-05;  
Matches 172; Conservative 75; Mismatches 304; Indels 227; Gaps 36;

Qy 43 VTGLARGGLPVAMEVVAALHAPLDAF-----VVRKLGAPGHDFEFAVGA-----LAS 88  
Db 269 VTSLSPEFMFPAALSGAGGVH-PLRLADPGLAVVQAAEDGGLAPCSKGRRCRP 327  
Qy 89 GGRVVVNDVVRGLRITPQQLRDIAEREGRELLRESAYRGERPPTDTGKTVIVDDGL 148  
Db 328 GSRLV-----TVRPMPSGEQAGVAVRGV-----VQAQFPGADADPILAVVDAGQ 374  
Qy 149 ---ATGASMPAAVQALRPAQF-----AQIVINVA----- 175  
Db 375 QLEALLAGNLAALAAVEDDVQPLDAQRAASTDQATLVQVAAPOVETDAALADQPALAHQ 434  
Qy 176 -----AP-----ESTCEFAGLVDVVCAITMPTPFLAVGESFMDPROVTEE 217  
Db 435 ALDQAQVAFPGDDPAAVGVADPACQF-----QFAVAGQAATVQVGAQ 480  
Qy 218 VRRLATPTAGPSLRPPAS---TAAD-----VLRRVADAPEGVPTHEVLAEIV-GDA 267  
Db 481 AQRLLATGGAATVVEGAAVQVHMLADQATLVGVQASGDAQVAGEQALAAVVGGA 540  
Qy 268 RIVLIGE---SSGTHGFYQARAAMTQMLIEKGFQVAA-----EADNPDAVY 314  
Db 541 -----GERQAGAGNGAFGVVQRAVDAQVDAAGQAGALGVQASGRDVQAGAGDHG 595  
Qy 315 NRYVRIGGEDT-----NADEALSGFERFPAMMMRNTVVVDPEVWLRTRNQRYESGALRQAG 370  
Db 596 QAVVQGLDADQAGQAGADQAPAAVAVQAGGEGEGAVADPFRAGAVV--HRLAIAQQQKAG 653  
Qy 371 FYGLDLYSLHRSIOEVSIVLDKVPRAAARARAYACPDHACADDDQAYGFAAAGAPS 430  
Db 654 -----RGDQAAVA-----VDQRAIIEVGC-----DAGFADQDASALVEAFQVGE-- 692  
Qy 431 CERAEVQVLDVQRRNALAYARQDGLLADELFTYAQQNAQTVRDAAEVYTRAMPSGRVTSWN 490  
Db 693 -----VEQALGADSLILAVQPG-----HGGDAGVADAADAIVAIVQAHAG----- 732  
Qy 491 LRDMHMTGLSLILTH-----LDRH--LDAAPARIYV-----WAINSHVGDARA 532  
Db 733 -ADTH--RITGA--DHAGTAVVEAGALQRIHAGIEQRIALVVQRLAGQORRTGAGEGPA 787  
Qy 533 TEVVAQGLTLTGQIVRER-----YGDERSRISGFSTYTGTVTAASEWGIQORKAVERPAL 586  
Db 788 TVVQARBARQAAAFADQAAALVVQAAABAHQAQVLAVERPTAVAVQFAAV-QAQAQVAAQ 846  
Qy 587 H--GSVEELFHQIADSF-----LVSAKLSRDAEAPLDVVRIGRAIG-VYUUPATERQSHYL 639  
Db 847 HPLGLVQOALHGEQAQAAVADDLAAAVVOLLAGVHGDLRGAGNLAGAIVVDLPRLSDAAR 906

Qy 640 HVRP-----ADQFAMIHIDQTALEBLEVTSRWIAGENPETVPTGL 681  
Db 907 GDQGLAVVDRVGRDLQGLFADQFATLLGQAARQLQVA-----LGGDTFSGVADGV 958

RESULT 6  
US-09-902-540-15988  
; Sequence 15988, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(115849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; PRIORITY FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 15988  
; LENGTH: 2214  
; TYPE: PRT  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-15988

Query Match 4.1%; Score 143.5; DB 2; Length 2214;  
Best Local Similarity 23.1%; Pred. No. 0.00016;  
Matches 173; Conservative 80; Mismatches 302; Indels 193; Gaps 37;

Qy 5 AADVTRSPRRVPRDRREAGRVLAELAAVYRDP-DVIVLGLARGGLPVAMEVVAALHA 63  
Db 1278 ATAELVYRRADLFDDKDA---IDALLPLARLRPAASVIDRA-----VAGLHA 1324  
Qy 64 PLDAFVVRKLGAPGH--DEPAVGLASGGRVVNDVVRGLRITPQQLRDIAEREGRELL 121  
Db 1325 -----LGRHDLIDVVAAGAAAGRRAA-ELLAAASVASSST--ADEDAWTL 1371  
Qy 122 RRESAYRGERPPTDTGKTVIVDDGLATGASMPAAVQALRDAQVIVAVPAPBSTC 181  
Db 1372 TORAA---EAAPELDTALQALVT--GLRQGDSTRLEALE---QWPRVEDADEAV 1421  
Qy 182 REF-----AGVDDVVCATMPTPFLAVGESFMDPROVTEBEVRLATPTAGPSLR 233  
Db 1422 LRLELAELADAGVD-----VARELDEVARAGSAGAG-- 1456  
Qy 234 PAASTADVLRRAVIDAPGVPTHEVL-AEIV-GDARIVLIGSSHGTHGFYQARAAMTQ 291  
Db 1457 ---ADALEALEPLLKQAPARAAYQVABAELASGRBOVLMAAARG---FESAGQLPE 1509  
Qy 292 WLIEKGFQAVAAADPMDAYRVNRVYRGLGEDTNADEALSGFERFPAMMMRNTVVVDPEV 351  
Db 1510 ALKAAR--DAASVPDVALRVNHLTVRASGEAPRAARALLQAAARLAPEBRPPLLE-- 1565  
Qy 352 EWLRTRNQRYESGALRQAGFYGLDLYSLHRSIOEVSIVLDKVPRAAARARAYACPDHA 411  
Db 1566 -----AAGMEKXAGHGEALVEFERIATA---SDVLABSELNERGRIGAFARA 1612  
Qy 412 CADDDQAYGFAPAAAGAPSGSCEAEVQVLDVQRRNALAYARQD-----LTAE 458  
Db 1613 L-----EVGFAPAAASG-----DLTD--ALMAAQAGDTARTREALMALALAD 1654  
Qy 459 DELFYAQONACTV---DAE--VYTRAMPSGRVTSW--NLRDQMAQTLSLTHLDRHL 511  
Db 1655 ADPAAHAAALDGLRABDABGLELELAGISAAKDAAFVVALRDEVLRSAASVLRRL-RAL 1713  
Qy 512 D--APP-----ARIYVMAHNSHVGDAR-----ATEVVAQGLT 542  
Db 1714 EELAPFEGFAARLTLLPLTEKLPBALAEAVLTQVNAQPGTVREALMAADGFPSSRKS 1773  
Qy 543 LGQIVRERYGDERSRISGFSTYTGTVTAASEWGIQORKAVERPALHGSVEEL---FHQTA 598



Db 1774 L---LRRRYALE---LRLGRYACVRLTSLQVAGESSDAVAAVLAHLRGGILLSLPLRLP 1827  
QY 599 DSFLVSAKRLSDAPAE-----LDVYRLGALGVYTLPATERSHYLAH-----RP 643  
Db 1828 EAEAEAFQAKDPAQQLSLRHLSTLVLDGEE-PAVEFSLVERLEQAAGVSGNTYERL 1886  
QY 644 ADQFDAM-IHIDPRALEPLEVTSRWTA 670  
Db 1887 ADAYEALGVADAQAQLERLPEIDERLA 1914

## RESULT 7

US-09-252-991A-26723  
; Sequence 26723, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-07-27  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26723  
; LENGTH: 685  
; TYPE: PRN  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26723

Query Match 3.9%; Score 136; DB 2; Length 685;  
Best Local Similarity 24.9%; Pred. No. 0.00013;

Matches 133; Conservative 57; Mismatches 117; Indels 164; Gaps 34;

QY 9 VTRRSRRVRDRRRAAGVLAELIAAYRDQPDVVLG-LARGGLPVAME-----VAAAL 61  
Db 108 VQRR--RRFAGAPRQAGROLAHA-----RGAVDVVDVGTVAESAQFAYGGGQGGMAADL 161  
QY 62 HAPLDAFVVRKLGAP-----GHDRFAVGALASGGRVNVDDVVRGRTTPQOLR-DIA 113  
Db 162 E---EAVVQAEVGMPOHFLRGGHLOFAAGHA-----LIVESFLALVVQVSTQGLAVDLA 214  
QY 114 ERERRELLRRRESAY-----RGERPPTDITGKTVIVVDG-----LATGA- 152  
Db 215 VDRGRQAVQRAVAGQHVVGQFLAQCGLDGSEHGR-----GIAVLADHDHEGAQLLATGAM 270  
QY 153 ---SMPRAVQALRDAPQAVIIVAPRAPESTCEEPAGLVDDVVCATMPPTPFLAVGESFW 208  
Db 271 VGRKHHRRAVDS-RLAQQRGLDVA-----EFRAVADLHLVVQPT----- 309  
QY 209 DFRQVNDDEVRRLATPTAGPSLRPAASTADVLRVAIDAPGS-VPTHEVLAEIVGDA 267  
Db 310 -----EELQALIGQRA---GLVAGAVEPTPOCLR--VEAPRGHVRVAQVFPAR-QADA 356  
QY 268 RIVLIGSSHGHTHEFYQARAAMTQWLIEKGFAGVAALAD---WPDAYRVVRVY---RG 320  
Db 357 RQVQL-----TSQVAAARPVVG---VEDEGLAVIDRADRGSQPMCFGVDDAAGSDRG 407  
QY 321 LGEEDTNADEL---SGFER---FRAMWRTVVRDFTVEMLRTNRQRYESGALRQAGFYGL 374  
Db 408 LGRAVVLVHNGIGCPACCDRQFFPG-----AEQVAQR-QRRRGGLLRGRH--- 452  
QY 375 DIVSLHSIGSEVSYLDKVDPRAAARARAYACFDH-----ACADGGQAY--GFAA 423  
Db 453 ---HNGRREGVA---DPLIGEECQRRRVAAHLVQGVQEAQAQGGPDPFLGNVE 502  
QY 424 A---FGAGPSCREBA-----VEQLVDVQRNALAYARODGLIADDEL 461  
Db 503 ADAGHOGGAARPLIHIEARVVRPAHEVGGATLDHNLALGFA--GGRRGEDHI 550

## RESULT 8

US-07-689-008-2  
; Sequence 2, Application US/07689008  
; Patent No. 5268274  
; GENERAL INFORMATION:  
; APPLICANT: Ben-Bassat, Arle  
; APPLICANT: Calhoun, Roger D  
; APPLICANT: Fear, Anna L  
; APPLICANT: Gelfand, David H  
; APPLICANT: Meade, James H  
; APPLICANT: Tal, Rony  
; APPLICANT: Wong, Hing  
; APPLICANT: Benizman, Moshe  
; TITLE OF INVENTION: METHODS AND NUCLEIC ACID SEQUENCES FOR THE  
; TITLE OF INVENTION: EXPRESSION OF CELLULOSE SYNTHASE OPERON  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McCutchen, Doyle, Brown & Eversen  
; STREET: Three Embarcadero Center  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/689,008  
; FILING DATE: 19910422  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 337,194  
; FILING DATE: 12-APR-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 496,236  
; FILING DATE: 23-MAR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy, Lisabeth Felix  
; REGISTRATION NUMBER: 31547  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 393-2000  
; TELEFAX: (415) 393-2286  
; TELEX: 340817 MACPAG SFO  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3031 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-689-008-2

Query Match 3.8%; Score 134.5; DB 1; Length 3031;  
Best Local Similarity 19.5%; Pred. No. 0.0023;

Matches 152; Conservative 96; Mismatches 298; Indels 235; Gaps 34;

QY 16 RVFRDRRRAAGVLAELI---LAAYRQPDVIVGLARGGLPVAMEVAAALHAPLDAF 68  
Db 1465 QILAQKKNQAKIGDVLNAGDDLTSYRSSPLYTV-----GYPLWLEPVMVHNNPSRV 1519  
QY 69 VVRKLAGHDEFAVGALASGGRVNVDDVVRGLRTTPQQLDIAREGELLRRRESAYR 128  
Db 1520 IV-----VGLL---GCILIVAVVVRAL-----AGHALRRRRELQ 1550  
QY 129 GERPPPTDITGKTVIVDDGLATGASMPRAVQALRDAPQAVIIVAPAA---PESTCREFA 185  
Db 1551 EERQRT-NNRRYVLSIGALASSCM-----TVLAVVAPARAQASTAMTTA 1596  
QY 186 GLVDVVCATMPPTPFLAVGESFMDFRQVTDDEVVRLLATPTAGPSL----- 231  
Db 1597 A---TSATPAARQIILLQARFVLQDQOQYDMMQALQNNERLAPNSPBDVLEVGIVOTAI 1652

Qy 232 -RRPAASTADVLRRVAIDAPGVPT---HEVLAE-----LV 264  
Db 1653 GNRRA---ADTLRLHQVAFCGSAAGNLNDLSRAISQSDLSQIRSLAGSGQAQVA 1709  
Qy 265 GDARIVLIGESSHG--THEFYQARAM--TOMLIEEGFCAVAADWDPAIRVN--RIVRGL 321  
Db 1710 GYOKLPHGKPKPHSLAVEYQTMAGVPAQWDAKAGLGVVASN--PODYEAQALFAQAL 1767  
Qy 322 GEDTN-----ADBALSGFERPPAMMMRTYVVRDPE--WLRTOR- 357  
Db 1768 TYNSTRMEGLTRLKDLOSFRSQAEVEAAAAQSTRQTLSDLPVNPETQPLMEQHLNHP 1827  
Qy 358 -----NORYESGALRQAGFYGLD---LYSLHRSIOEVIY-----L 390  
Db 1828 NDTALREHMLHPGSPRPDKAGLARQAGYQQLNAGSLAAAEQSPQALQINSHDLSLGM 1887  
Qy 391 DKVDPRAAARARARYACFDHACADDGQAYG--FAAFAQAGSCREAVEQLV-----DV 442  
Db 1888 GLVSRQGDTEAR--RYFEEMAADPKTADBRPRLAGMAVSGEYASVROLIAHQYTEA 1946  
Qy 443 ORNALAVARODGLADELFFYA--QQNAQTVRDAEVVYRAMPSGRVTSWMLRDQMAQTL 500  
Db 1947 KQQLATLARQPGQYTGATMLADLQKSTGQIAAEQETRGILS-----REPNNQAL 1998  
Qy 501 GSLTLHLDRHLDAPPARIVVAHNSHVGDARATEVMADGQTLGQIVRERYGDSRS--IG 559  
Db 1999 MGL-----ARVDM-----ACQNTAEARQLLSRVPQYASQVG 2030  
Qy 560 FSTYGTIVTAASEWGGIAQRAVRPALHGSVEELFHQTAQSFVLSARLSRAEAPLDVVR 619  
Db 2031 EIEVSGMAAASQSDSARKVSI-----LREMAQAQRPDPWVYINLANLQCGQDVAE 2083  
Qy 620 LGRALGVVY---LATERQSHYLVHRPADQPDAMIHQTRALRELVTSRWINGENET 676  
Db 2084 AGRWQPLANPVAQDRQAGILTYTSGSN--DAM-----TROILAGLSPAD 2128  
Qy 677 Y 677  
Db 2129 Y 2129

RESULT 9  
US-09-679-279-14  
; Sequence 14, Application US/09679279  
; Patent No. 6524841  
; GENERAL INFORMATION:  
; APPLICANT: McDaniel, Robert  
; APPLICANT: Volchegursky, Yanina  
; TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic  
; TITLE OF INVENTION: Genes and Uses Thereof  
; FILE REFERENCE: 300622004700  
; CURRENT APPLICATION NUMBER: US/09/679, 279  
; CURRENT FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/158,305  
; PRIOR FILING DATE: 1999-10-08  
; PRIOR APPLICATION NUMBER: US 60/190,024  
; PRIOR FILING DATE: 2000-03-17  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FaastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 3562  
; TYPE: PRT  
; ORGANISM: Micromonospora megalomicea  
US-09-679-279-14

Query Match 3.8%; Score 134; Length 3562;  
Best Local Similarity 24.1%; Pred. No. 0.0034;  
Matches 172; Conservative 64; Mismatches 273; Indels 204; Gaps 40;

Qy 47 ARGGLPVAMEVAALAHAPLDAFVVRKIGAP--GHDFAVGALASGGRVVVND--VVRGLR 103  
Db 2580 APGKLPFAMR--GVTLHA--TGATAVAVVATPAGPDAVALRVTDPTGQLVATVDALVVRDAG 2637

Qy 104 ITPQQLDIAREGRELRLRESAYRGERPTDITGKTVIVVDGLATGASWFAAVQALRD 163  
Db 2638 ADROQPRG---RDS--DLHRLIEWVRLATPPDT--PAVVHVAADGLD-----DLRLA 2682  
Qy 164 AQPQIVIAV---PAAESTCRERAGLV-----DDVCAIMPTEPLAVGESFWD 209  
Db 2683 GGPAQAVVVRVYRQDGPDAEARNHGVLMATVLRWMLDD--RWPATTLVVAATG-- 2735  
Qy 210 FROYTDEVRRLATPTAGPSLRPPASTADVLT-----RVAID-----APGVPT 256  
Db 2736 ---AGVEV-----SPGDDVPRFGAAVWGVLRCQAQSPDRFVLVDGDPETPRPAVPD 2784  
Qy 257 HEVLAEIVGARIV-----LIGESSHGTHEFYQARAAMTOMLIEEGFCAV--AAEDW 308  
Db 2785 NPQALA--VRQGAIVVPRLTLAGVPVAVADRAYR-----LVPNG--SIEAVAPAV 2833  
Qy 309 PDAYR-----VNRVYRGLGEDTNAEDALGGERFPAMMMRTYVVRDPEVEMLRNRQYE 362  
Db 2834 PDADRPLAPBEVRVAVRATG--VNRFDVLALGVYRPAEWGTEASGVTVESGVRRFT 2891  
Qy 363 SGALRQA-----GFYGLDYSLHRSIOEVIYSDKVDPRPA--ARARARYACFDHACAD 414  
Db 2892 PG---QAVTGLFQCAFEPVAVADHRLTVPDPCKRAVDAALVPIAFTTAHYALHDLAQLQ 2948  
Qy 415 DGQA--YGFMAAFGAGPS---CEREAVE-----QLVDQRNALAVARODG-- 454  
Db 2949 AGQSVLHAAGAAGVMAAVALARRAGAEVFATASPAHPTLRALGLDDHILASRESGFG 3008  
Qy 455 -----LAEDELFFYAQOQACTVDAEYTYAMSGR 485  
Db 3009 ERPAARTGCGVDVVLNSLTGDLDESARLLADGVE--VEMGKTDLPAE-----QRRGR 3062  
Qy 486 VTSWNLDDQMAQTLGSLTH-----LDR-----HLDAPPARIVVAHNSHVDA 530  
Db 3063 YVFPDLAEAG--PDLRLGILSEVGLAAGALDRLPVSWELISAPALYTHMSRGRHVGL 3121  
Qy 531 RATE---VMADGQL-----TLGQIVRERYGDSRSIGSTYGTIVTAASEWG---GI 576  
Db 3122 VLTQPAVHPDGTVLVVGSTIGRLV-----ARHLVYGHGVPHLLVASRRRPAAPGA 3174  
Qy 577 AQRRAVRPALHGSVEELFHQTAQSFVLSARL--SHDAEAPLDVVRLGRALGVY 628  
Db 3175 AELRADVEGIGATIEIVACDTADREALAALDLSIPADRL-----TGVVH 3219

RESULT 10  
US-09-252-991A-17140  
; Sequence 17140, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252, 991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 17140  
; LENGTH: 535  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17140

Query Match 3.7%; Score 131; DB 2; Length 535;  
Best Local Similarity 25.4%; Pred. No. 0.00028;  
Matches 137; Conservative 44; Mismatches 193; Indels 166; Gaps 29;

Qy 3 MTAADVTRSPRRVFRDREAGRVLAEILAAVYRDQDVIIVLGIARGLPVAME--VAAAL 61

Db 36 LPAAGGARRRRPRVADLRRLA--ARRLARIRPLPRLAALLRAGAGGRRMGLAAMP 93  
Qy 62 HAPL-----DAFVVRKLGAPGDEFAVAGALASGGRVVNDVVGRLITPQOLRDIAERE 116  
Db 94 HVALAGSAAALERRRL-----RLAVGR-----ERRRGLR--PLRGHAPR 135  
Qy 117 GRELLRESAYRGERPPTDTGKTVIVVDGLATGASMFPAVALRDAQAQIVY----- 171  
Db 136 GR--ARR--VQPRRLPADRVGHRL-----GQPGANDLVGGRSLPGADPARPALRGKR 185  
Qy 172 -AVPAARESTCRERAGLVDDVVCATMP-----TPFLAVGESFNFQVTEDEVRLLATP 225  
Db 186 GTGVAPAAARRARA-----DV--ASLPRRRPRRVPL-----RQLRRRPLP 225  
Qy 226 -----TAGSLRRPASTAADVLRRAVD-----APGVPTHEVLA 261  
Db 226 LRQDGLRAPGGLAESSRRRAAGLRRTLAAGRTAAALLGAGVSLASVVDHVALG 285  
Qy 262 E-----LVGDARIVLIGESSHGTHEFYQARAAVT-----OMLIEKGFQAV-- 302  
Db 286 HAAAGEDVAFGDILVDPBAVVLHLHLAGLH--YLAGAADAHPAGLRNVQAEIGGGVHG 343  
Qy 303 --AAEAD---WPAAYV--NRVVRGIGEDTNADEALSGFERPPAMMRRVTVDVEM 353  
Db 344 FPGADADALAGGGEFHLVLDGRYLRRLG-----GF----- 374  
Qy 354 LRTNRQRESGA-----LRQAGFYGLDLSLHRSIOEYISYDKVDPRAAR---ADARY 405  
Db 375 ---RGQPPVDGADLRRRRPFGLVVAHVVAVALQVVL-HRAHGRALAAEDPFLVEVRG 430  
Qy 406 ACPDHACADGQAY--GFAAFAFGAPSCERAEVQLVDVGRNALA--YARQDGLAADEL 461  
Db 431 DQFGHAGVQVATVAGPFAAVALADVDVQAEVVDPRGHVQLAVVDVLRRTGAVDEDHV 490

RESULT 11  
US-08-426-630-49  
Sequence 49, Application US/08426630  
Patent No. 6656709  
GENERAL INFORMATION:  
APPLICANT: BLANCHE, FRANCIS; CAMERON, BEATRICE; CROUZET,  
APPLICANT: JOEL; DEBUSCHE, LAURENT; LEVCY SCHIL, SOPHIE;  
APPLICANT: THIBAUT, DENIS  
TITLE OF INVENTION: POLYPEPTIDES INVOLVED IN THE  
TITLE OF INVENTION: BIOSYNTHESIS OF COBALAMINS AND/OR COBALAMIDES, DNA SEQUENCES  
TITLE OF INVENTION: CODING FOR THESE POLYPEPTIDES, PREPARATION METHOD AND THEIR  
TITLE OF INVENTION: USE.  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 555 13TH STREET, N.W.  
CITY: WASHINGTON  
STATE: DISTRICT OF COLUMBIA  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,630  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/916,151  
FILING DATE: 14-SEP-1992  
APPLICATION NUMBER: PCT/FR91/00054  
FILING DATE: 30-JAN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: F. F. CALVETTI  
REGISTRATION NUMBER: 28,557

REFERENCE/DOCKET NUMBER: 1290-7213  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 857-7887  
TELEFAX: (202) 857-7929  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1275 amino acids  
TYPE: Amino acid  
STRANDEDNESS:  
TOPOLOGY: Linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: No  
ORIGINAL SOURCE:  
ORGANISM: Pseudomonas denitrificans  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
HAPLOTYPE:  
TISSUE TYPE:  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
FEATURE:  
NAME/KEY: COBN  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION: Translation product of SEQ ID NO:48  
US-08-426-630-49

Query Match 3.7%; Score 131; DB 2; Length 1275;  
Best Local Similarity 20.0%; Fred. No. 0.0012;  
Matches 127; Conservative 87; Mismatches 235; Indels 186; Gaps 29;

Qy 65 LPAFVVRKLGAPGDEFAVAGALASGGRVVNDVVGRLITPQOLRDIAEREGRELLRE 124  
Db 683 LPAVLDLKEMOJRDGHTFGVAPBGRLLT--DITVALARVPRGLGEGGDSIQRAIAAD 740  
Qy 125 SAYRGERPPTDTGKTVIVVDGLATGASMFPAVALRDAQ-----AQ 168  
Db 741 AGLRGAIFTSAGNDPA-----RDAQPPDPDPCVMSDTWTGPKPS 780  
Qy 169 IYAVAAARESTCRERAGLVDDVVCATMPFPLAVGESFNFQVTEDEVRLLATPTAG 228  
Db 781 ILADLDAPWRITGDVVERI-ELIANTLVSGELACPDHANTAVVGEIETRL-----K 833  
Qy 229 PSLRPAASTAADVLR-----RVAIDAPGVPTHEVLAELVGDARIVLIGESSHGTHEFY- 283  
Db 834 PSISNGAEMTGFLTGLSGRFVAPGPGCAPTR-----GRPDVLPTRGN-----FTS 880  
Qy 284 -QARAAMTOMLIEKGFVAE-----ADWPDAYRVNRV-----VRGIGEDTNAD 328  
Db 881 VDSRAVPTPAAYE--LGKSAELLIRRYLDHGEWPSFGLAWGTAMRTGGDDIAQA 937  
Qy 329 EALSGERPPAMMRRVTVDVFEMLRTNQRKE-----SGALRQAGFGDL 376  
Db 938 LALIGAK--PTW-----DWSRRVMEIYIPLAVLGRPRVDVLRISGFF-RDA 983  
Qy 377 YS-----LHRSIOEYISYDKVDPRAAARARAYACFDHACDDGAYGPA--ARGAAP 429  
Db 984 FPDQIALFDPAIRAVALEEDDADNMIAARRRARSRLBEGVEAARARASTRVGAKP 1043  
Qy 430 SCERAEVQLVD-----VORNALAYARQDGLAADELFYAQONQVTR--- 472  
Db 1044 GAYGALLQALIDKEKMETKADLAELVTWCAVAYVAGGEGSKARDLF--EERRTIEAVV 1101  
Qy 473 -----DAVYYRAMFSGRVTSKMLRDQHAQTLGSLTLHLDRHLDAAPPAIYVM 521  
Db 1102 QNODNREHLDLSDDYQ--FEGGMSA-----AAEQLGHRPAIYHNDHSRPERVIR 1152  
Qy 522 AHSNHYD-ARATEV--WADQGLTGQIVRERYGDSISIGSTYGTVTASE-----W 573  
Db 1153 SLEEEIGRVRAVVPKIDGV-----RHG-----YKGAPEIATVDYMF 1194

Qy 574 GGIAQRKAVRPAHSGVEELFHOADSFLVSARLS 608  
Db 1195 AFAATTGAVR-----DHHEFAAYQAFIVDERVA 1222

## RESULT 12

US-09-410-551B-72  
Sequence 72, Application US/09410551B  
Patent No. 6503737  
GENERAL INFORMATION:  
APPLICANT: KOSAN BIOSCIENCES, Inc.  
APPLICANT: REEVES, CHRISTOPHER  
APPLICANT: CHU, DANIEL  
APPLICANT: KHOSLA, CHAITAN  
APPLICANT: SANTI, DANIEL  
APPLICANT: WU, KAI  
TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
TITLE OF INVENTION: CONSTRUCTS THEREFOR  
FILE REFERENCE: 30062-20026.00  
CURRENT APPLICATION NUMBER: US/09/410,551B  
PRIOR FILING DATE: 1999-10-01  
PRIOR APPLICATION NUMBER: US 60/139,650  
PRIOR FILING DATE: 1999-06-17  
PRIOR APPLICATION NUMBER: US 60/123,810  
PRIOR FILING DATE: 1999-03-11  
PRIOR APPLICATION NUMBER: US 60/102,748  
PRIOR FILING DATE: 1998-10-02  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 72  
LENGTH: 6396  
TYPE: PRT  
ORGANISM: Streptomyces hygroscopicus  
US-09-410-551B-72

Query Match 3.7%; Score 130; DB 2; Length 6396;  
Best Local Similarity 21.0%; Pred. No. 0.024;  
Matches 163; Conservative 64; Mismatches 274; Indels 274; Gaps 31;

Qy 17 VFRRRREAGR-----VLAELLAAYRD-----QPDVIVLGLARGG----- 50  
Db 4453 MFRGMRAMWDGDTVAEVALPEBRADADGFGNHPALDAAQSGSLMLLESQESVQ 4512  
Qy 51 LPVAMEVAALAHAPLDAFVVRKLGAPGHDEFAVGALASGGRVVVNDVVRGLRTTPOOLR 110  
Db 4513 LPFSWM-GVRFHA-TGATMLRVAVVPDGLRLHAADSNGRPVATIDAL-----VTRSPEA 4566  
Qy 111 DIAR-----EGRELLRR--ESAYRGERP 132  
Db 4567 DLAPADMLRVGMAPVPVPAAGPSDADVLTLRGDDADPLGETRDLTTRVLDALLRADRP 4626  
Qy 133 PTDTGKTIVVDDGLATGSMFAAYQALRDAPQAVIAVPAAPESICREFAGLVDDV 192  
Db 4627 -----VIRQVVTGLAKAKAA--GLVRTAQNEQPGREFL-VETDP-----GEVL 4666  
Qy 193 CATMPTPLAVGSEFMDFRQVTDSEVRRLATPT-----AGPSLRPPASTAADVLRRVA 247  
Db 4667 DGAKRDAIALGEPHYVLRDGLFPAARLMKATPSPILTPDGTSGWQDRSATSGLDVLAVP 4726  
Qy 248 IDAB-----GGVPTHEVLAEL--VGDA-----IVLIGESSHGTHF 282  
Db 4727 TDAPDRPLAGEVRIAVRAAGLNFPRDVTVALGVVADARPLGSBAAGVVL--ETGPGVHDL 4784  
Qy 283 YQAPAAATQWILIEBKGCAGAVAAEADWPAYRVNRYVKGLEDTYADELSGFERF-PAWM 341  
Db 4785 APGRVL-----GMLAGAFGPVALITDRRLIGMPDGTFFPQAASVMTAFATAM- 4832  
Qy 342 WRNTVVDPEVEMLTRNQRYESGALRQAGFYGL-DVYSIHSIOEVISYLDKVDPPRAAR 400  
Db 4833 -----VIRQVVTGLAKAKAA--GLVRTAQNEQPGREFL-VETDP-----GEVL 4666  
Qy 401 ARAR-----YAC-----FDHACADDGQAYGPAAR----- 425

Db 4862 QIANHLGAERYATTSAAKRHLVLDLGDGHLADSSTATAADAPPPVDVVLNSITGEFLDASV 4921  
Qy 426 -----GAGPSCREAVEQLVDV-----ORNALAYARODGLAEDELFPYAQNAQT 470  
Db 4922 GLLAAGRPTEMKGTDIRHAVQGPFDLMDAGPRMQRIIYBGLFPAVDVLHPLPVAMWD 4981  
Qy 471 VRDAEVYTRAMFSGRVTSMNLRODHNAQTGSLTLTHDRHLDAFPARIIVVMAHNSHVDA 530  
Db 4982 VQARREAFGMWSSGRHT-----GKLVLTVPRLD-PEGAVVI----- 5017  
Qy 531 RATEVMAHQGLTLCQIYRERYGDE-----SRISGFSITYGT----- 566  
Db 5018 -----TGSGGTLAGIILARHLGHPHYLLSRTPPTTPGTHLPDGVDPHQAATTIARI 5071  
Qy 567 ---VTAASEWGIAQRKAVRPAHSGVEELFHOADSFLVSARLSRDAEALVDV 618  
Db 5072 PQLITAVFHTAGTLDLALNLTFRDVRDVLTKPRADAAWMLHRLTRDTDLAFAFV 5126

RESULT 13  
US-09-940-316B-72  
Sequence 72, Application US/09940316B  
Patent No. 6759536  
GENERAL INFORMATION:  
APPLICANT: KOSAN BIOSCIENCES, Inc.  
APPLICANT: REEVES, CHRISTOPHER  
APPLICANT: CHU, DANIEL  
APPLICANT: KHOSLA, CHAITAN  
APPLICANT: SANTI, DANIEL  
APPLICANT: WU, KAI  
TITLE OF INVENTION: POLYKETIDES ENCODING THE FKBA GENE OF THE FK-520 POLYKETIDE SYNTHASE  
TITLE OF INVENTION: GENE CLUSTER  
FILE REFERENCE: 30062-20026.11  
CURRENT APPLICATION NUMBER: US/09/940,316B  
CURRENT FILING DATE: 2001-08-27  
PRIOR APPLICATION NUMBER: 09/410,551  
PRIOR FILING DATE: 1999-10-01  
PRIOR APPLICATION NUMBER: US 60/139,650  
PRIOR FILING DATE: 1999-06-17  
PRIOR APPLICATION NUMBER: US 60/123,810  
PRIOR FILING DATE: 1999-03-11  
PRIOR APPLICATION NUMBER: US 60/102,748  
PRIOR FILING DATE: 1998-10-02  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 72  
LENGTH: 6396  
TYPE: PRT  
ORGANISM: Streptomyces hygroscopicus  
US-09-940-316B-72

Query Match 3.7%; Score 130; DB 2; Length 6396;  
Best Local Similarity 21.0%; Pred. No. 0.024;  
Matches 163; Conservative 64; Mismatches 274; Indels 274; Gaps 31;

Qy 17 VFRRRREAGR-----VLAELLAAYRD-----QPDVIVLGLARGG----- 50  
Db 4453 MFRGMRAMWDGDTVAEVALPEBRADADGFGNHPALDAAQSGSLMLLESQESVQ 4512  
Qy 51 LPVAMEVAALAHAPLDAFVVRKLGAPGHDEFAVGALASGGRVVVNDVVRGLRTTPOOLR 110  
Db 4513 LPFSWM-GVRFHA-TGATMLRVAVVPDGLRLHAADSNGRPVATIDAL-----VTRSPEA 4566  
Qy 111 DIAR-----EGRELLRR--ESAYRGERP 132  
Db 4567 DLAPADMLRVGMAPVPVPAAGPSDADVLTLRGDDADPLGETRDLTTRVLDALLRADRP 4626  
Qy 133 PTDTGKTIVVDDGLATGSMFAAYQALRDAPQAVIAVPAAPESICREFAGLVDDV 192  
Db 4627 -----VIRQVVTGLAKAKAA--GLVRTAQNEQPGREFL-VETDP-----GEVL 4666  
Qy 193 CATMPTPLAVGSEFMDFRQVTDSEVRRLATPT-----AGPSLRPPASTAADVLRRVA 247

Db 4667 DGAKRDALALGEBHVRLROGLFEARILMRATPSLTLPDTGSKQLRPSATGSLDDLAIVP 4726  
Qy 248 IDAP-----GGVPTHEVLAEL--VGDAE-----IVLIGESSHGTHF 282  
Db 4727 TDADDRPLAGEVRLAIVRAAGLNFREDYVALGVADAPRLGSEAAVVL--ETGPGVYDL 4784  
Qy 283 YQABAAATQMLIEERKGAVALAEDWPRAYRVNRVYKGLGSDTNADELSGFEF--PAM 341  
Db 4785 APGRVL-----GMLAGAFGPVATIDRRLGMPDGMTPPOAAVMTAFATW- 4832  
Qy 342 WRNTVVDVFEVMTLRNQRYESGALRQAGFYGL-DLYSLHRSIOGVISYLDKVPRAAR 400  
Db 4833 -----YGLVDLAGL-RPGKVLHHAATGVGAAY 4861  
Qy 401 ARAR-----YAC-----FDHACADGQAYGFAAF----- 425  
Db 4862 QIARHLAGAEVYATTSAAKRLVLDGAILADSRTAFADAPPVVDVNLSTGEFLDASV 4921  
Qy 426 -----GAGSCREAVEQLVDV-----QRNALAYARQDGLLADELFTYQONAO 470  
Db 4922 GLLAAGRFLEMGTDIRHAHQDPDLMDAGPDMORIYELLGLFADVLAHPLPVHAMD 4981  
Qy 471 VRDAEYVYRAMFGSRVTSWNLRDQMAQTGLSLTLTHDRHLDAPPARIVVWAHSHVGA 530  
Db 4982 VRQAREAFGMSSGRHT-----SKLVLTVPRLD--PEGAVI----- 5017  
Qy 531 RATEWMDGOLTLGQIVRERYGDE-----SRISGFSTYTG----- 566  
Db 5018 -----TGSGTGLAGILARHLGHPHTYLSRTPRPDTTGGTHLPCVDGPHQLATTLARI 5071  
Qy 567 ---VTAASEWGLQKRAVNPALHGSVELFHQADSELSVARSRAEARLDVY 618  
Db 5072 POPLAVHTAGTLDLALDNLTPDRVDTVLKPRADAMHLRLTRDTLAAFFV 5126

RESULT 14  
US-09-477-962-117  
; Sequence 117, Application US/09477962  
; Patent No. 6927286  
; GENERAL INFORMATION:  
; APPLICANT: SHEN, BEN  
; APPLICANT: DU, LIANGCHENG  
; APPLICANT: SANCHEZ, CESAR  
; APPLICANT: CHEN, MEI  
; APPLICANT: EDWARDS, DANIEL J.  
; TITLE OF INVENTION: BLEOMYCIN GENE CLUSTER COMPONENTS AND THEIR USES  
; FILE REFERENCE: 407T-895820US  
; CURRENT APPLICATION NUMBER: US/09/477,962  
; CURRENT FILING DATE: 2000-01-05  
; PRIOR APPLICATION NUMBER: 60/115,435  
; PRIOR FILING DATE: 1999-01-06  
; PRIOR APPLICATION NUMBER: 60/118,848  
; PRIOR FILING DATE: 1999-02-05  
; NUMBER OF SEQ ID NOS: 133  
; SOFTWARE: PatentIn Ver. 3.0  
; SEQ ID NO 117  
; LENGTH: 2841  
; TYPE: PRT  
; ORGANISM: Streptomyces verticillius  
; FEATURE:  
; OTHER INFORMATION: ORF32  
US-09-477-962-117

Query Match 3.7%; Score 128.5; DB 2; Length 2841;  
Best Local Similarity 20.9%; Pred. No. 0.0086;  
Matches 158; Conservative 70; Mismatches 263; Indels 265; Gaps 36;

Qy 4 TAAADYTRR-----SPRRVFRDRREAGRVLAE-LTAAVYDQPDVTVLGIAGGLP----- 52  
Db 1966 TAAPVVITRLRLPADPGVHIDPRQARDPRDLRLHRDQAAVITFTSGSGAPKGVVVR 2025  
Qy 53 -----VAMEVAAALHAPLDAFVVRKUGAPQH-----DEFA 82

Db 2026 HGSLYHLGHVRRMAEGCPRRANVAHTTAMTFDDSLSEFLMLVAGHTLHVAPEVRNDPEA 2085  
Qy 83 VGALASGGRVAVDDVVRGLRITPQQLRDIAREBEGHELLRRESAYNGERPPTDITGKTVI 142  
Db 2086 LVALLVRRAIDV-----LNVTPSHL-----TLTIEAGLIEGDRVP-----GTVL 2124  
Qy 143 VVDDGL-----ATGA-----SMEPAVQALRDA-----QPAQVIVAPPAPESTCR 182  
Db 2125 VGBEAVPALMRLTLREKTREFENLYGPTAEVTDACHDLSPADVPVLTPLPHTHVR 2184  
Qy 183 EFAGLVDDVVCATMPPTFLVAGESF----- 207  
Db 2185 -----VLD-----RLRPVVGAGELIYLGCTGLARGYLNRPALYQRFVADPYPTPSRLY 2237  
Qy 208 -----W-----DFRQVTDSEVR-----RLIAPYAGSLRRPAASTAAD--VL 243  
Db 2238 RTGDRARWRBDGTLEYLGRTRDDQIKIRGFVPEGEIEAVLTHHRAVEAAVTVATDGA 2297  
Qy 244 RRVALDAPGVPTHEVLAELVGDARVYLIGESSHGHEFYQADAAATQMLIEERGF----- 229  
Db 2298 RLVALVAPARAPHGSADGAPDAQV-----EENNAVFEATHTDAADGELTFENIKGWD 2353  
Qy 300 -GA-VAAE--ADMPD-----AYRVNRVYRGEGEDTNADEALSGFERPAMMR-- 343  
Db 2354 TGAPIRAEHRKREVDITTVARLLERPARVLEIGSGTG-----LTMKRL 2397  
Qy 344 -----NTVVDVFEVMTLRNQRYESGALR-----QAGFYGL-----DLYSLHRSIO 385  
Db 2398 PHTEYTGTFSPRAVDMLDGLRRRPARHRVRLHREATDFTGVRAASTDLVVNVSVQY 2457  
Qy 386 V--ISYLDKVPRAARARARACFDHACDDG-----QAYGFAAGAGP 429  
Db 2458 FPDRAVLDTVLARAL-----DATADRGRVFGVDVRNIALAPQFAARQALAHAGP 2506  
Qy 430 SCREAVEQLVDQVRNALAYARQDGLAEDELFTYAOONQOTVDAEYVYRAMFGSRVTSW 489  
Db 2507 GAAR-----DARAAGEAAMDGELVSPAYFALAAAR-----SPRTGV 2547  
Qy 490 NL--RDOH--MAQTGLSLTLH--DRHLDAPPARIVV--AHNSHVGADARATEVMAD 538  
Db 2548 EILPRGRHNEHNSLVRYDVVLHVGGDRPA-ADEAEVLTGWDQVHDLASLSARLGRGPD 2606  
Qy 539 GQLTGQIVRERYGDSRSIGFSTYGTVTAASEWG 574  
Db 2607 ALLVRG-VANDRLTRDNELDAPARTTAVPEPDLWG 2641

RESULT 15  
US-10-042-665A-8  
; Sequence 8, Application US/10042665A  
; Patent No. 6924106  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Toupet, Christine  
; APPLICANT: Engel, Nathalie  
; TITLE OF INVENTION: Rifamycin biosynthesis gene cluster  
; FILE REFERENCE: 4-21001/B/C1  
; CURRENT APPLICATION NUMBER: US/10/042,665A  
; CURRENT FILING DATE: 2002-01-09  
; PRIOR APPLICATION NUMBER: 09/242,744  
; PRIOR FILING DATE: 1999-03-24  
; PRIOR APPLICATION NUMBER: PCT/EP97/04495  
; PRIOR FILING DATE: 1997-08-18  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 3413  
; TYPE: PRT  
; ORGANISM: Amycolatopsis mediterranei  
US-10-042-665A-8

Query Match 3.7%; Score 128.5; DB 2; Length 3413;  
Best Local Similarity 21.3%; Pred. No. 0.012;



**THIS PAGE BLANK (USPTO)**

**This Page Blank (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using SW model

Run on: March 23, 2006, 06:14:33 ; Search time 143.21 Seconds  
(without alignments)  
1986.885 Million cell updates/sec

Title: US-10-617-038-23

Perfect score: 3508  
Sequence: 1 VLMTAAADVTRSRPRVRFRD.....LEVTSMIAGBNPTYPEPGL 681

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_RA\_Main:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3508	100.0	681	US-10-617-038-23	Sequence 23, Appl
2	431	12.3	443	US-11-045-942-10	Sequence 10, Appl
3	431	12.3	443	US-11-058-869-10	Sequence 10, Appl
4	431	12.3	443	US-11-110-578-10	Sequence 10, Appl
5	417.5	11.9	233	US-11-045-942-4	Sequence 4, Appl
6	417.5	11.9	233	US-11-058-869-4	Sequence 4, Appl
7	417.5	11.9	233	US-11-110-578-4	Sequence 4, Appl
8	401.5	11.4	229	US-11-045-942-5	Sequence 5, Appl
9	401.5	11.4	229	US-11-058-869-5	Sequence 5, Appl
10	401.5	11.4	229	US-11-058-869-5	Sequence 5, Appl
11	401.5	11.4	229	US-11-058-869-5	Sequence 5, Appl
12	401.5	11.4	229	US-11-110-578-5	Sequence 5, Appl
13	401.5	11.4	229	US-11-110-578-5	Sequence 5, Appl
14	389	11.1	210	US-10-156-761-13240	Sequence 13240, A
15	389	11.1	210	US-10-156-761-13240	Sequence 13240, A
16	179	5.1	179	US-10-156-761-7635	Sequence 7635, Ap
17	157	4.5	236	US-09-881-752A-340	Sequence 340, App
18	152	4.3	234	US-10-335-977-9382	Sequence 9382, Ap
19	149	4.2	2552	US-10-282-122A-61672	Sequence 61672, A
20	147	4.2	996	US-10-282-122A-6165	Sequence 996, A
21	145	4.1	921	US-10-156-761-9301	Sequence 9301, Ap
22	145	4.1	3192	US-10-203-295-6	Sequence 6, Appl
23	144	4.1	1741	US-10-282-122A-50287	Sequence 50287, A
24	143	4.1	836	US-10-282-122A-69284	Sequence 69284, A
25	143	4.1	836	US-10-132-134-34	Sequence 34, Appl
26	141	4.0	7338	US-10-211-028-11	Sequence 11, Appl
27	140	4.0	168	US-10-369-493-9033	Sequence 9033, Ap

28	136.5	3.9	1340	US-10-282-122A-47780	Sequence 47780, A
29	136	3.9	574	US-10-324-967-36	Sequence 36, Appl
30	136	3.9	846	US-10-369-493-9102	Sequence 9102, Ap
31	136	3.9	3753	US-09-980-217-29	Sequence 29, Appl
32	136	3.9	7066	US-10-203-295-20	Sequence 20, Appl
33	136	3.9	9477	US-10-203-295-37	Sequence 37, Appl
34	136	3.9	10625	US-10-819-386A-3	Sequence 3, Appl
35	134.5	3.8	730	US-10-437-963-193818	Sequence 193818, A
36	134	3.8	8438	US-10-937-730A-4	Sequence 4, Appl
37	133.5	3.8	1126	US-10-369-493-11877	Sequence 11877, A
38	133.5	3.8	5065	US-10-171-311-52	Sequence 52, Appl
39	133.5	3.8	5065	US-10-408-765A-1743	Sequence 1743, Ap
40	133.5	3.8	5065	US-10-631-467-758	Sequence 758, App
41	133	3.8	1743	US-10-819-386A-2	Sequence 2, Appl
42	132.5	3.8	889	US-10-852-707-109	Sequence 109, App
43	131.5	3.7	1745	US-10-314-657-35	Sequence 35, Appl
44	131.5	3.7	1745	US-10-473-193-35	Sequence 35, Appl
45	130	3.7	6396	US-09-940-316B-72	Sequence 72, Appl

## ALIGNMENTS

```
RESULT 1
US-10-617-038-23
; Sequence 23, Application US/10617038
; Publication No. US20040057963A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Peter
; APPLICANT: Rosenkrands, Ida
; APPLICANT: Strjhn, Anette
; TITLE OF INVENTION: Therapeutic TB Vaccine
; FILE REFERENCE: SSIUSAUSA
; CURRENT APPLICATION NUMBER: US/10/617, 038
; CURRENT FILING DATE: 2003-07-11
; PRIOR FILING DATE: 2002-07-13
; PRIOR APPLICATION NUMBER: DK PA 2002 01098
; PRIOR FILING DATE: 2002-07-13
; PRIOR APPLICATION NUMBER: US 60/401, 725
; PRIOR FILING DATE: 2002-08-07
; NUMBER OF SEQ ID NOS: 187
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 681
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-10-617-038-23
```

```
Query Match 100.0%; Score 3508; DB 4; Length 681;
Best Local Similarity 100.0%; Pred. No. 4.8e-305;
Matches 681; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 VLMTAAADVTRSRPRVRFRDREGRVLAELAAVROPVIVIGLARGGLPVAMVEVAAA 60
D 1 VLMTAAADVTRSRPRVRFRDREGRVLAELAAVROPVIVIGLARGGLPVAMVEVAAA 60
QY 1 LHAFLDFFVVRKLGAPGHDFEFAVGAALASGRVNVNDVVGRLTTPQOLRDIAEREGREL 120
D 1 LHAFLDFFVVRKLGAPGHDFEFAVGAALASGRVNVNDVVGRLTTPQOLRDIAEREGREL 120
QY 61 LHAFLDFFVVRKLGAPGHDFEFAVGAALASGRVNVNDVVGRLTTPQOLRDIAEREGREL 120
D 61 LHAFLDFFVVRKLGAPGHDFEFAVGAALASGRVNVNDVVGRLTTPQOLRDIAEREGREL 120
QY 121 LRRSAYRGRPPPTDITGKTIVVDDGLATGASVFAVQALRDAQPAQIVAVPAPEST 180
D 121 LRRSAYRGRPPPTDITGKTIVVDDGLATGASVFAVQALRDAQPAQIVAVPAPEST 180
QY 121 LRRSAYRGRPPPTDITGKTIVVDDGLATGASVFAVQALRDAQPAQIVAVPAPEST 180
D 121 LRRSAYRGRPPPTDITGKTIVVDDGLATGASVFAVQALRDAQPAQIVAVPAPEST 180
QY 181 CREPAGLVDDVVCATMTPLPVLAVGESFWDROYTDEVRLLATPTGSPSLRPAASTAA 240
D 181 CREPAGLVDDVVCATMTPLPVLAVGESFWDROYTDEVRLLATPTGSPSLRPAASTAA 240
QY 241 DVLRRVVIDAPGVPVTHVLAELVGDARIVLIGSSSHGTHFEYQARAAMTOMLIEEGCFG 300
D 241 DVLRRVVIDAPGVPVTHVLAELVGDARIVLIGSSSHGTHFEYQARAAMTOMLIEEGCFG 300
QY 301 AVAAEADWPAYRVNRYVVGIGEDTNADEALSGFERFPAMMMRNTVVRDFEVLRTNOR 360
D 301 AVAAEADWPAYRVNRYVVGIGEDTNADEALSGFERFPAMMMRNTVVRDFEVLRTNOR 360
```



Db 301 AWAADWPAAYRVNRVYRGIGEDTNADEALSGFERPPAMMNTVVRDVEMLRTNOR 360  
Qy 361 YEGSALHQAGFYGLDLSLHRSIQEVIISYLDKVPRAAARAAAYACGDHACADGQAYG 420  
Db 361 YEGSALHQAGFYGLDLSLHRSIQEVIISYLDKVPRAAARAAAYACGDHACADGQAYG 420  
Qy 421 FAAAAGAGPSCEBAVEQVLDVQORNAALAYAQDGLAEDELFYAQONNAQTVRDAEYVYRA 480  
Db 421 FAAAAGAGPSCEBAVEQVLDVQORNAALAYAQDGLAEDELFYAQONNAQTVRDAEYVYRA 480  
Qy 481 MFSGRVTSWNLRDQHMOTGSLTLHLDRHLDAPARIVVAHNSHGDARATVWADQ 540  
Db 481 MFSGRVTSWNLRDQHMOTGSLTLHLDRHLDAPARIVVAHNSHGDARATVWADQ 540  
Qy 541 LTTGQIVRERYGDSRSIGSFYTGTVVAASEMGIAQKAKVAPALHSGVELFHQTADS 600  
Db 541 LTTGQIVRERYGDSRSIGSFYTGTVVAASEMGIAQKAKVAPALHSGVELFHQTADS 600  
Qy 601 FLVSARLSRDAEAPLDVVRGRLAIGVYVLPATERQSHYLVHVPADQPDAMTHIQDTALE 660  
Db 601 FLVSARLSRDAEAPLDVVRGRLAIGVYVLPATERQSHYLVHVPADQPDAMTHIQDTALE 660  
Qy 661 PLEVTSRWIAGENBETYPGTL 681  
Db 661 PLEVTSRWIAGENBETYPGTL 681

## RESULT 2

US-11-045-942-10  
; Sequence 10, Application US/11045942  
; Publication No. US20050172350A1  
; GENERAL INFORMATION:  
; APPLICANT: Kloeck, Andrew P.  
; APPLICANT: Williams, Deryck Jeremy  
; APPLICANT: Salmon, Brian  
; TITLE OF INVENTION: NEMATODE PPPT-LIKE SEQUENCES  
; FILE REFERENCE: 12557-005001  
; CURRENT APPLICATION NUMBER: US/11/045,942  
; CURRENT FILING DATE: 2005-01-27  
; PRIOR APPLICATION NUMBER: US/10/113,201  
; PRIOR FILING DATE: 2002-03-29  
; PRIOR APPLICATION NUMBER: US 60/280,192  
; PRIOR FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 443  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis PPPT  
US-11-045-942-10

Query Match 12.3%; Score 431; DB 6; Length 443;  
Best Local Similarity 32.4%; Pred. No. 1,8e-29;  
Matches 129; Conservative 52; Mismatches 161; Indels 56; Gaps 11;

Qy 16 RVFRDRREAGRVLAELAAAYRDQPDVIVGLARGGLPVAMEVAAALHAPLDAFVVRKLG 75  
Db 2 KLFDRDGRDAGRLAQLAQLSGKA-VVVLGLPRGVPVPAFEVAKSLQAPLDVLVVRKLG 60  
Qy 76 PGHDEPFAVGALASGRVYVNDVVRGLRITPQQLRDIAEREGRELLRESAYGERPPTD 135  
Db 61 PQPBLAFGAIGBDGVVLNDVVRGTHLDAAMDVERKQILQLQRAERFRGRDRIP 120  
Qy 136 ITGKTIVVDDGLATGASMPAAVOALRDAQPAQIVIVPAAPESCTGEPAGLVDDVCAT 195  
Db 121 LTGRIAVIVDDGLATGATATAACQVAAHAGADKVLAVIPGPDIVARFAGYDEVVCLA 180  
Qy 196 MPTPFLAVGSGFMDPQVQTVDEVRRLATPTAGPSLRPAASTADVLRRVAIDAGGVP 255  
Db 181 TPALFFAVGGGYRFTQTSDEVAFL-----DRAHRPFAEKALDAADDP 227  
Qy 256 THEVLAEVGDARIVLIGESSHGTHERYQARAANTQWLIEKFGAIVA-AEADWPDAYRV 314  
Db 256 THEVLAEVGDARIVLIGESSHGTHERYQARAANTQWLIEKFGAIVA-AEADWPDAYRV 314

Db 228 LRDEEVUVAGP-----VPVAGHLT-----VPEKRGIVVPAHSGSSRHSI 269  
Qy 315 -NRVY-----RGLEDVNADEALSGFER-----FPAMMNTVVRDVEMLRTNORYES 363  
Db 270 RNRVYAEVLVAGAPATLLFLLTPBEERBRANVFDEILASRLI-DVTGLATQ-----PD 324  
Qy 364 GALKRQAGFYGLDLSLHRSIQEVIISYLDKVPRAAARAAAYACGDHACADGQAYG 420  
Db 325 TASLPVGYFG-----ASTGAGALVVAADPFRVNVRA 355

## RESULT 3

US-11-058-869-10  
; Sequence 10, Application US/11058869  
; Publication No. US20050172363A1  
; GENERAL INFORMATION:  
; APPLICANT: Kloeck, Andrew P.  
; APPLICANT: Williams, Deryck Jeremy  
; APPLICANT: Salmon, Brian  
; TITLE OF INVENTION: NEMATODE PPPT-LIKE SEQUENCES  
; FILE REFERENCE: 12557-005001  
; CURRENT APPLICATION NUMBER: US/11/058,869  
; CURRENT FILING DATE: 2005-02-16  
; PRIOR APPLICATION NUMBER: US/10/113,201  
; PRIOR FILING DATE: 2002-03-29  
; PRIOR APPLICATION NUMBER: US 60/280,192  
; PRIOR FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 443  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis PPPT  
US-11-058-869-10

Query Match 12.3%; Score 431; DB 6; Length 443;  
Best Local Similarity 32.4%; Pred. No. 1,8e-29;  
Matches 129; Conservative 52; Mismatches 161; Indels 56; Gaps 11;

Qy 16 RVFRDRREAGRVLAELAAAYRDQPDVIVGLARGGLPVAMEVAAALHAPLDAFVVRKLG 75  
Db 2 KLFDRDGRDAGRLAQLAQLSGKA-VVVLGLPRGVPVPAFEVAKSLQAPLDVLVVRKLG 60  
Qy 76 PGHDEPFAVGALASGRVYVNDVVRGLRITPQQLRDIAEREGRELLRESAYGERPPTD 135  
Db 61 PQPBLAFGAIGBDGVVLNDVVRGTHLDAAMDVERKQILQLQRAERFRGRDRIP 120  
Qy 136 ITGKTIVVDDGLATGASMPAAVOALRDAQPAQIVIVPAAPESCTGEPAGLVDDVCAT 195  
Db 121 LTGRIAVIVDDGLATGATATAACQVAAHAGADKVLAVIPGPDIVARFAGYDEVVCLA 180  
Qy 196 MPTPFLAVGSGFMDPQVQTVDEVRRLATPTAGPSLRPAASTADVLRRVAIDAGGVP 255  
Db 181 TPALFFAVGGGYRFTQTSDEVAFL-----DRAHRPFAEKALDAADDP 227  
Qy 256 THEVLAEVGDARIVLIGESSHGTHERYQARAANTQWLIEKFGAIVA-AEADWPDAYRV 314  
Db 228 LRDEEVUVAGP-----VPVAGHLT-----VPEKRGIVVPAHSGSSRHSI 269  
Qy 315 -NRVY-----RGLEDVNADEALSGFER-----FPAMMNTVVRDVEMLRTNORYES 363  
Db 270 RNRVYAEVLVAGAPATLLFLLTPBEERBRANVFDEILASRLI-DVTGLATQ-----PD 324  
Qy 364 GALKRQAGFYGLDLSLHRSIQEVIISYLDKVPRAAARAAAYACGDHACADGQAYG 420  
Db 325 TASLPVGYFG-----ASTGAGALVVAADPFRVNVRA 355

## RESULT 4

US-11-110-578-10  
; Sequence 10, Application US/11110578  
; Publication No. US20050186631A1  
; GENERAL INFORMATION:

```

1  APPLICANT: Kloeck, Andrew P.
2  APPLICANT: Williams, Deryck Jeremy
3  APPLICANT: Salmon, Brian
4  TITLE OF INVENTION: NEMATODE PPPT-LIKE SEQUENCES
5  FILE REFERENCE: 12557-005001
6  CURRENT APPLICATION NUMBER: US/11/110,578
7  CURRENT FILING DATE: 2005-04-20
8  PRIOR APPLICATION NUMBER: US/10/113,201
9  PRIOR FILING DATE: 2002-03-29
10 PRIOR APPLICATION NUMBER: US 60/280,192
11 PRIOR FILING DATE: 2001-03-30
12 NUMBER OF SEQ. ID NOS: 10
13 SOFTWARE: PasteSeq for Windows Version 4.0
14 SEQ. ID NO 10
15     LENGTH: 443
16     TYPE: PRT
17     ORGANISM: Mycobacterium tuberculosis PPPT
18     US-11-110-578-10

```

Query Match	12.3%;	Score 431;	DB 6;	Length 443;
Best Local Similarity	-32.4%;	Pred. No. 1.8e-29;		
Matches 129;	Conservative 52;	Mismatches 161;	Indels 56;	Gaps 11;

QY 16 RVRDRERGRVLAETLLAAYRQDPVITVGLRGLPAAWEEVAALPAIDAFVRLTGA 75  
 Db 2 KLFDDGDGGRQLAQRLADLSGKA-VVYIGLPRGVPAFEVAKSLQAPLDIVYRLTGV 60  
 QY 76 PGHDEPAVALASGGRVVNDVVDVGRITTPQOLRIAREGREILRESAYRGERPTD 135  
 Db 61 PFQPELAFGALIGEDGVRVINDVDVVRGTHLDAAMDAVERKQILIELQRAERFRGRDRIP 120  
 QY 136 ITSKTIVIVDDGLTASGMPFAAVQALRDAQPAQIYIANPAAPESTCEPAGLVDDVYCAT 195  
 Db 121 LTRIRIVIVDDGIALGATGATKAAACQVAPRAGADKVLAVIPGIDDIIVARFAGYADEVVICLA 180  
 QY 196 MPTPFLAVGESFMDFQVTVDEEVRRLTAPTAGPSLRPAASTADVLRLVAIDAPGVP 255  
 Db 181 TPALPFAVAGCGYRNFQTSDDDEVAVFL-----DRAHRDPRAGALIDAAADPP 227  
 QY 256 THEVLAEVLGDARIYVLTIGESSHGTHEFYQARAAMTOMLIEKSGGAVA-AEADWPDAYRV 314  
 Db 228 LRDEBTVQVAVP-----VPVAGHLT-----VPEKPRGIIVFPHGSSSSHSI 269  
 QY 315 -NRYV---RGEGEDTNADEALSGER-----PAMKMRNTVVRDFVEMVLTNRQRES 363  
 Db 270 RNRVYAEVLVTGAFATLLFDLLTPBEBERRANVPFIELIASRLI-DVYGMILATQ---PD 324  
 QY 364 GALRQAGFYGLDLYSLHRSIGEVISVLDKVRPRAARA 401  
 Db 325 TSLPVGIFG-----ASTGAGALVAAADPRVVNRA 355

```

RESULT 5 555-942-4
; Sequence 4, Application US/11045942
; Publication NO. US20050172350A1
; GENERAL INFORMATION:
; APPLICANT: Kloeck, Andrew P.
; APPLICANT: Williams, Deryck Jeremy
; APPLICANT: Salmon, Brian
; TITLE OF INVENTION: NEMATODE PEPT- LIKE SEQUENCES
; FILE REFERENCE: 12557-005001
; CURRENT APPLICATION NUMBER: US/11/045,942
; CURRENT FILING DATE: 2005-01-27
; PRIOR APPLICATION NUMBER: US/10/113,201
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: US 60/280,192
; PRIOR FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 233
; TYPE: PRT

```

US-11-045-942-4

Query Match	11.9%	Score 417.5;	DB 6;	Length 233;
Best Local Similarity	41.9%;	Pred. No. 1.1e-28;		
Matches 88;	Conservative 44;	Mismatches 75;	Indels 3;	Gaps 2;

QY 71 RKLGAHGDEFAVGALASG--RVVNDVDVYRGRLITPQQLRIADREGRELLRRSAYR 128  
 DQ 65 RKTGAHGHEEYGGIGAVNEGNPPELWMNEDAVKTYQTPREGYVQAMMERKQLTEIARQRRL 124  
 QY 129 GEPRPDDITGKVIYVDDGCLATGASMPAAVQLADDAQAQIVIAVPAAPESTCREFGLV 168  
 DQ 125 GDRPPLSLAAGKLAIVDDGSIATGAGTARVAMALRQKNVAKALLASPLAPSDTLALERAG 164  
 QY 189 DDVVCATMPPTPLAVGSEFMDPRQVTDDEV 218  
 DQ 185 NEVLVLETPPNFSAVGLHTKFDQDSDEV 214

```

RESULT 6
US-11-058-869--4
Sequence 4, Application US/11058869
Publication No. US20050172368A1
GENERAL INFORMATION:
APPLICANT: Kloeck, Andrew P.
APPLICANT: Williams, Deryck Jeremy
APPLICANT: Salmon, Brian
TITLE OF INVENTION: NEMATODE PPPT-LIKE SEQUENCES
FILE REFERENCE: 12557-005001
CURRENT APPLICATION NUMBER: US/11/058,869
CURRENT FILING DATE: 2005-02-16
PRIOR APPLICATION NUMBER: US/10/113,201
PRIOR FILING DATE: 2002-03-29
PRIOR APPLICATION NUMBER: US 60/280,192
PRIOR FILING DATE: 2001-03-30
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 233
TYPE: PRT
ORGANISM: Meloidogyne incognita PPPT
US-11-058-869--4

```

Query Match	11.9%;	Score 417.5;	DB 6;	Length 233;
Best Local Similarity	41.9%;	Pred. No. 1.1e-28;		
Matches 88;	Conservative 44;	Mismatches 75;	Indels 3;	Gaps 2;

[illegible]

RESULT 7  
US-11-110-578-4  
; Sequence 4, Application US/11110578  
; Publication No. US20050186631A1

```

; GENERAL INFORMATION:
; APPLICANT: Kloeck, Andrew P.
; APPLICANT: Williams, Deryck Jeremy
; APPLICANT: Salmon, Brian
; TITLE OF INVENTION: NEMATODE PPPT-LIKE SEQUENCES
; FILE REFERENCE: 12557-005001
; CURRENT APPLICATION NUMBER: US/11/110,578
; PRIOR FILING DATE: 2005-04-20
; PRIOR APPLICATION NUMBER: US/10/113,201
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: US 60/280,192
; PRIOR FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Meloidogyne incognita PPPT
US-11-110-578-4

```

```

Query Match 11.4%; Score 417.5; DB 6; Length 233;
Best Local Similarity 41.9%; Pred. No. 1.1e-26;
Matches 88; Conservative 44; Mismatches 75; Indels 3; Gaps 2;

```

```

Qy 12 RSPRRVRRRGRVLAELAAVYRDPD-VITGLARGLPVAMEVAAALHAPLDAFV 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 RAATAPPKDHDADQCKLAELKFKSQDKVVLALPRGGVPVAFEVAKSIGAPLDLIMV 64

Qy 71 RKAGPCHDEFAVAGALSGG--RVVNDVVRGRLITPQQLDIAERGRELLRESAYR 128
   ||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 65 RKIGAPGHEEYGAIVVGNPPELVNMDAVKTYQPEGVQAMKEQLKSTIAQRKRYL 124

Qy 129 GERPPITITGKTVIVDDGLATGASMPFAVQALRDAOPAQIVIAVPAPESTCREPAGLV 188
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 GDRRPLSLAGLAIIVDDGATGCTARVAMKALQKQVAKALLASPLASPTLAELEAEG 184

Qy 189 DDVVCATMPTPEFLAVGESFMDFRQVTDEEV 218
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 185 NEVLVLETPTNPFSAVGLHYTKFDQTSDEEV 214

```

```

RESULT 8
US-11-045-942-5
; Sequence 5, Application US/11045942
; Publication No. US20050172350A1
; GENERAL INFORMATION:
; APPLICANT: Kloeck, Andrew P.
; APPLICANT: Williams, Deryck Jeremy
; APPLICANT: Salmon, Brian
; TITLE OF INVENTION: NEMATODE PPPT-LIKE SEQUENCES
; FILE REFERENCE: 12557-005001
; CURRENT APPLICATION NUMBER: US/11/045,942
; CURRENT FILING DATE: 2005-01-27
; PRIOR APPLICATION NUMBER: US/10/113,201
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: US 60/280,192
; PRIOR FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Meloidogyne javanica PPPT
US-11-045-942-5

```

```

Query Match 11.4%; Score 401.5; DB 6; Length 229;
Best Local Similarity 40.2%; Pred. No. 3e-27;
Matches 84; Conservative 45; Mismatches 77; Indels 3; Gaps 2;

Qy 17 VFDRRBRAGRVLAELAAVYRDPD-VITGLARGLPVAMEVAAALHAPLDAFVRLTGA 75
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 9 LFIDRKDAGGKLAALAHILIPQRDNIVLALPRGGVPVACEAADAQAPLDLIMVKRTGA 68

```

```

Qy 76 PGHDEFVAGALASGG--RVVNDVVRGLITPQQLDIAERGRELLRESAYRGRRP 133
   ||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 69 PGHEEYGAIVVGNPPELVNMDAVKTYQPEGVQAMKEQLKSTIAQRKRYL 128

Qy 134 TDITGKTVIVDDGLATGASMPFAVQALRDAOPAQIVIAVPAPESTCREPAGLVDDVC 193
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 129 VSLGRIAIIVDDGATGCTARVAMKALQKQVAKALLASPLASPTLAELEAEGVLC 188

Qy 194 ATMPTPEFLAVGESFMDFRQVTDEEVRL 222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189 LETPPTNPFSAVGLHYQRFDQTSDEEVIRCM 217

```

```

RESULT 9
US-11-045-942-6
; Sequence 6, Application US/11045942
; Publication No. US20050172350A1
; GENERAL INFORMATION:
; APPLICANT: Kloeck, Andrew P.
; APPLICANT: Williams, Deryck Jeremy
; APPLICANT: Salmon, Brian
; TITLE OF INVENTION: NEMATODE PPPT-LIKE SEQUENCES
; FILE REFERENCE: 12557-005001
; CURRENT APPLICATION NUMBER: US/11/045,942
; CURRENT FILING DATE: 2005-01-27
; PRIOR APPLICATION NUMBER: US/10/113,201
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: US 60/280,192
; PRIOR FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Heterodera glycines PPPT
US-11-045-942-6

```

```

Query Match 11.4%; Score 401.5; DB 6; Length 229;
Best Local Similarity 40.2%; Pred. No. 3e-27;
Matches 84; Conservative 45; Mismatches 77; Indels 3; Gaps 2;

```

```

Qy 17 VFDRRBRAGRVLAELAAVYRDPD-VITGLARGLPVAMEVAAALHAPLDAFVRLTGA 75
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 9 LFIDRKDAGGKLAALAHILIPQRDNIVLALPRGGVPVACEAADAQAPLDLIMVKRTGA 68

Qy 76 PGHDEFVAGALASGG--RVVNDVVRGRLITPQQLDIAERGRELLRESAYRGRRP 133
   ||:|||||:|:~:~:~:~:~:~:~:~:~:~:~:~:~:~:~:~:~:~:~:~:~:~:~:~:~:
Db 69 PGHEEYGAIVVGNPPELVNMDAVKTYQPEGVQAMKEQLKSTIAQRKRYL 128

Qy 134 TDITGKTVIVDDGLATGASMPFAVQALRDAOPAQIVIAVPAPESTCREPAGLVDDVC 193
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 129 VSLGRIAIIVDDGATGCTARVAMKALQKQVAKALLASPLASPTLAELEAEGVLC 188

Qy 194 ATMPTPEFLAVGESFMDFRQVTDEEVRL 222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189 LETPPTNPFSAVGLHYQRFDQTSDEEVIRCM 217

```

```

RESULT 10
US-11-058-869-5
; Sequence 5, Application US/11058869
; Publication No. US20050172363A1
; GENERAL INFORMATION:
; APPLICANT: Kloeck, Andrew P.
; APPLICANT: Williams, Deryck Jeremy
; APPLICANT: Salmon, Brian
; TITLE OF INVENTION: NEMATODE PPPT-LIKE SEQUENCES
; FILE REFERENCE: 12557-005001
; CURRENT APPLICATION NUMBER: US/11/058,869
; CURRENT FILING DATE: 2005-02-16
; PRIOR APPLICATION NUMBER: US/10/113,201
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: US 60/280,192

```

```

: PRIOR FILING DATE: 2001-03-30
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 5
: LENGTH: 229
: TYPE: PRT
: ORGANISM: Meloidogyne javanica PPPT
: US-11-058-869-5

```

Query Match 11.4%; Score 401.5; DB 6; Length 229;  
Best Local Similarity 40.2%; Pred. No. 3e-27;  
Matches 84; Conservative 45; Mismatches 77; Indels 3; Gaps 2;

[illegible]

RESULT 11  
US-11-058-869-6  
; Sequence 6, Application US/11058869  
; Publication No. US20050172363A1

Query Match	11.4%;	Score 401.5;	DB 6;	Length 229;
Best Local Similarity	40.2%;	Pred. No. 3e-27;		
Matches 84;	Conservative 45;	Mismatches 77;	Indels 3;	Gaps 2;

[illegible]

RESULT 12  
US-11-110-578-5  
; Sequence 5, Application US/11110578  
; Publication No. US20050186631A1

QY 17 FRPRRREGRVLAETLLAAYRRQPD - VITLGLARGLPVAEVAALAHMLDLAFVRKKGDA 75  
Db 9 LFRIRKDGQKIAKLALAHILPQRDNI VVALAPRGVPAVACEADAPQALDILMRKIGA 68  
QY 76 PGHDEFAYGALASGG - RVVVVNDVDVVRGRTIPQOOLRIABREGRELLRESAYRERPP 133  
Db 69 PGHEHYGIGAVVEGNPPBELVMNNEEAVRTTRPEEAYVQOQMQLKEMEQRTKTYLGDKPP 128  
QY 134 TDIITGKTYIVDDGLATGASMFAYAOALRDAQPAQIVIAVPAPESTCFREAGLYDDVVC 193  
Db 129 VSLBERIAIVDDGATGSTARVAMKALRQKNVSAILLASPVAAPDYLALAEAGNEVLC 188  
QY 194 ATMPPLAVAGESFMWDFRQVTDEEVRLL 222  
Db 189 LETPPNFSAVGILHYQRFDTSDSEVTRCGM 217

RESULT 13  
US-11-110-578-6

; Sequence 6, Application US/11110578  
; Publication No. US20050186631A1

APPLICANT: Kloeck, Andrew P.

```

1 FILE NAME: 041101, 01 0411
2
3 TITLE OF INVENTION: NEMATODE PPPT-LIKE SEQUENCES
4
5 FILE REFERENCE: 12557-005001
6

```

Query Match	11.4%	Score 401.5;	DB 6;	Length 229;
Best Local Similarity	40.2%	Pred. No. 3e-27;		

	Matches	84,	Conservative	45,	Mismatches	77,	Indels	3,	Gaps	2.
Qy	17	VDRDRREGVLAETLLAAYRDQP-D-VIVLGARGGLPVAMEVAALAHAPLDAFVVRKIGA	75	:	:	:	:	:	:	:
Dd	9	LFIIRKDGGQTKARKLAHLIPQRDNIWVLALPRGCVPAACEAADDFQAPLDILMRKTIGA	68	:	:	:	:	:	:	:
Qy	76	PGHDEFAGTALSGS-RVVNVDDVVRGRITPQQLRIAEERREGLLRRESAYNGSRPP	133	:	:	:	:	:	:	:
Dd	69	PGHEBYTGAVAGEGNPPPELVMMEEAVRITTRSEAAVYQQMQKQLEMEREÖRTTYIGDXKP	128	:	:	:	:	:	:	:
Qy	134	TDITGKTVIYVDGDTASGMFAAQVALRDAQPAQIVIAVPAPBESTCREFGADVDPVC	193	:	:	:	:	:	:	:
Dd	129	VSLEGRIAIYVDDGIATGTARVALKALKLRQKNVSAILASPAPSDTLAELRAEGNEVLC	188	:	:	:	:	:	:	:
Qy	194	ATMPPTPLAVEGEFWDFRQVTDEEVRILL	222	:	:	:	:	:	:	:
Dd	189	LETPEPNFSAGLHYORFDQTSDEEVRCQM	217	:	:	:	:	:	:	:

RESULT 14  
US-10-156-761-13240  
; Sequence 13240, Application US/10156761  
; Publication No. US20030119018A1

```

1  APPLICANT: OMURA, SATOSHI
2  APPLICANT: IKEDA, HARUO
3  APPLICANT: ISHIKAWA, JUN
4  APPLICANT: HORIKAWA, HIROSHI
5  APPLICANT: SHIBA, TADAYOSHI
6  APPLICANT: SAKAKI, YOSHIYUKI
7  APPLICANT: HATTORI, MASAHIRA
8  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
9  FILE REFERENCE: 249-262
10 CURRENT APPLICATION NUMBER: US/10/156,761
11 CURRENT FILING DATE: 2002-05-29
12 PRIOR APPLICATION NUMBER: JP 2001-204089
13 PRIOR FILING DATE: 2001-05-30
14 PRIOR APPLICATION NUMBER: JP 2001-272697
15 PRIOR FILING DATE: 2001-08-02
16 NUMBER OF SEQ ID NOS: 15109
17 SEQ ID NO 13240
18 LENGTH: 210
19 TYPE: PRT
20 ORGANISM: Streptomyces avermectinib
21 US-10-156-761-13240

```

Query Match	11.1%	Score 389	DB 4	Length 210
Best Local Similarly	46.6%	Pred. No. 3.4e-26		
Matches 96	Conservative 27	Mismatches 75	Indels 8	Gaps 4

[illegible]

RESULT 15  
US-10-156-761-14961  
; Sequence 14961, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI

```

; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-252
; CURRENT APPLICATION NUMBER: US/0/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14961
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14961

Query Match          7.4%; Score 260; DB 4; Length 481;
Best Local Similarity 24.4%; Pred. No. 4.3e-14;
Matches 130; Conservative 66; Mismatches 236; Indels 100; Gaps 23

```

```

; ORGANISM: Streptomyces avermitilis
US-10-156-761-14961

```

[illegible]

Search completed: March 23, 2006, 06:28:11  
Job time : 144.21 secs



Db 1257 GALLPRLGRLVAPIRATGCGVEDESSLGRMAAFVAREBQVLDTVQTAVTLIGH 1316  
Qy 193 -CATMPTFLVAGSEFWDFROVTDEVR-----RLLATP----- 225  
Db 1317 GTATRAVDPAITFAETGFD--SLTAVELNRLRTATGRLSATLFDYTPRAVLGHLLRE 1314  
Qy 226 ---TAGSLRRPAASTAAD-----VLRRVAIDAPGCGPTEHEVLALVGDARIVLIGSSH 277  
Db 1375 LDGTVGAEVTRPAPAAATDRDPYIVGMACRYPGVAPSELEWELLATGR----- 1425  
Qy 278 GTHFYGARAMTQWLEKGFGAVAABDMPDAVRVRYRGGED--TNADEALSGFE 335  
Db 1426 -----DAYADLPDD-----RGMDLDELVSADPDSSG-- 1451  
Qy 336 RFPAMMMNTVVRDFVEWLRTRNORYESGALROAGFYG-----LDLYSLHRSIQEVL-S 388  
Db 1452 -----TSYVRSGGFYDAGEF-DADFGISPRALAMPDQGRLLLEVAME 1495  
Qy 389 YLDVNDPRAARARACPDHACADGQAYGPAAGFAGPSCREAVEQLV---DYQR 444  
Db 1496 TVERAGVPAASLKSQTCGVFVGAAA--QGYG---TGAGQAAGSGSEGYFLTGGAGSVVS 1548  
Qy 445 NALAY-----ARODGLLADELFEYAQNAQTVRDAEVYRRAMPFSGRYTSMNLRDQMA 497  
Db 1549 GRLSYTTEGEPRAVYVDTACSSSLVALHLAQLRSGEC--SLALAGVT-----VM 1598  
Qy 498 QTLGSLTLHLDRHLDAFPARIVVAMNSHVGDAATEWADGQTLGQIVERYGDESR- 556  
Db 1599 ATPGIFVEFSRQGLAADGRCKAP-----DAADGTGMSG--VGMLLERLSIDARRN 1649  
Qy 557 -----SIGFSTYTGTVTYAASEWGIQKRAVRPAL-----HGSV 590  
Db 1650 GHRVLAVRGSVAVNDGASNGELTAPN--GPSQQRVIRALANAGLAASDVDAVEALGTG 1706  
Qy 591 BELFHQADSFLVARSDAEAPL-----DYVRLGRATGVILPA 631  
Db 1707 TSL-GDPEAGALATYGGQEREPPLGLSISNIGHTOSAAAGVACIVMLVAMHGAFLPR 1765  
Qy 632 TEROSHYLVAPADQFDMIHID-QTRALEPLEVTSRKIAGENP 674  
Db 1766 T-----LHV---DQ--PSTHVDWSAGAVELTTERPAEHPGTSRP 1798

RESULT 2  
US-10-724-598-49  
Sequence 49, Application US/10724598  
Publication No. US20060019352A1  
GENERAL INFORMATION:  
APPLICANT: BLANCHE, FRANCIS; CAMERON, BEATRICE; CROUZET,  
JOEL; DEBUSSCHE, LAURENT; LEVCY SCHIL, SOPHIE;  
THIBAULT, DENIS  
TITLE OF INVENTION: POLYPEPTIDES INVOLVED IN THE  
BIOSYNTHESIS OF COBALAMINS AND/OR COBAMIDES, DNA SEQUENCES  
CODING FOR THESE POLYPEPTIDES, PREPARATION METHOD AND THEIR  
USE.  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & PINNEGAN  
STREET: 555 13TH STREET, N.W.  
CITY: WASHINGTON  
STATE: DISTRICT OF COLUMBIA  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10724,598  
FILING DATE: 01-Dec-2003  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/916,151  
FILING DATE: 14-SEP-1992  
APPLICATION NUMBER: PCT/FR91/00054  
FILING DATE: 30-JAN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: F. P. CALVETTI  
REGISTRATION NUMBER: 28,557  
REFERENCE/DOCKET NUMBER: 1290-7213  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 857-7887  
TELEFAX: (202) 857-7929  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1275 amino acids  
TYPE: Amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: Linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Pseudomonas denitrificans  
STRAIN: <Unknown>  
INDIVIDUAL ISOLATE: <Unknown>  
DEVELOPMENTAL STAGE: <Unknown>  
HAPLOTYPE: <Unknown>  
TISSUE TYPE: <Unknown>  
CELL TYPE: <Unknown>  
ORGANELLE: <Unknown>  
FEATURE:  
NAME/KEY: COBN  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION: Translation product of SEQ ID NO: 48  
SEQUENCE DESCRIPTION: SEQ ID NO: 49:  
US-10-724-598-49

Query Match 3.7%; Score 131; DB 6; Length 1275;  
Best Local Similarity 20.0%; Pred. No. 0.11;  
Matches 127; Conservative 87; Mismatches 235; Indels 186; Gaps 29;

Qy 65 LDAFVVRKLGAPGHDRAVALASGGRVYVNDVVRGRLTPQQLMDIAREBERELLRE 124  
Db 683 LDAYLCLDKEMQIRIDGLHIFGVAPEGRLT--DLTVALARVPRGLBGGQSLQRAIAD 740  
Qy 125 SAYGERPPTDITGKTYIVVDGLATGASWFAVQLRDAQP-----AQ 168  
Db 741 AGURRAIPTSAGGNRA-----RDAQPPDPLDCWSDTWGPKPS 780  
Qy 169 IVIAVPAPESTGREFAGLVDDVVCATMPTFLAVGESFWDFROVTDEVRLLATPTAG 228  
Db 781 ILADLSDAFRTAGDVTVERI-ELLANLVSGELACPDHMANTRAIVAGEIETRL-----K 833  
Qy 229 PSURRPAASTAAVLR---RVAIDAPGVPTHEVLALVGDARIVLIGSSHGTHEY- 283  
Db 834 PSISNGAAMETFLTGLSGRFVAPGSGAPTR-----GRPDVLPYTGNN-----FYS 880  
Qy 284 -QARAANTQWLEBEKGFVAVAE-----ADMPDAYVRVRY---VRGLGEPDNNAD 328  
Db 881 VDSRAVPTPAAYE---LGKKSABELLRRYIQDHGEWFPSSGGLTAKGTAMRTGGDDIAQA 937  
Qy 329 EALSGERFPAMMMNTVVRDFVEWLRTRNORYE-----SGALROAGFYGLDL 376  
Db 938 LALIGAK--PTW-----DWSRRVMGYEIVPLAVIGRPVVDVTLRISGFF-RDA 983  
Qy 377 YS-----LHRSIQEVSYLDKVDVPRRAARARAYACPDHACADGQAYGPA--AAGAGP 429  
Db 984 FPDQIALFDPAIRAVALBEDDADNMIAARRRARSRLAEVGVAABAAARRASRYVFGAKP 1043  
Qy 430 SCREAVEQLVD-----VORVALAYARODGLLADELFEYAQNAQTVR--- 472  
Db 1044 GAYGAALQALIDKGGMETRADLAAYVTWGAAYVYAGGEBKBERDLF--BERLRTIEAVV 1101



Qy 473 -----DAEYTYAAMSGRVTSWNLRDQHMAQTLSLTLTHLDRHLDAAPPARIYVW 521  
Db 1102 QONDRREHDLSDSDYQ--FEQGMSA-----AAEOLGGRPAIYHNDHSRPEKPYIR 1152  
Qy 522 AHNHVGD-ARATEV---WADGQLTGLGIVREYXGDESRSIGFSTYGTVAASE---W 573  
Db 1153 SLEBEIGVVRARVVRNPKRIDGM-----RHO-----YKCAFETIAATVDYMF 1194  
Qy 574 GGIAQRKAVRPALHSGVEELFHQTADSFLVSARLS 608  
Db 1195 AFAATGTGAVR-----DHFEAAVQAQATIVDERVA 1222

RESULT 3  
US-11-143-980-49  
; Sequence 49, Application US/11143980  
; Publication No. US20050272133A1  
; GENERAL INFORMATION:  
; APPLICANT: He, Min  
; APPLICANT: Hucul, John  
; APPLICANT: Haitli, Bradley A.  
; APPLICANT: Wegenaar, Melissa M.  
; APPLICANT: Graziani, Edmund  
; APPLICANT: Summers, Mia  
; APPLICANT: Kulowski, Kerry  
; APPLICANT: Pong, Kevin  
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Production of a Complex  
; TITLE OF INVENTION: Polyketide  
; FILE REFERENCE: AM-101426US  
; CURRENT APPLICATION NUMBER: US/11/143,980  
; CURRENT FILING DATE: 2005-06-03  
; PRIOR APPLICATION NUMBER: US 60/664,483  
; PRIOR FILING DATE: 2005-03-23  
; PRIOR APPLICATION NUMBER: US 60/576,895  
; PRIOR FILING DATE: 2004-06-03  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 49  
; LENGTH: 7968  
; TYPE: PR1  
; ORGANISM: Streptomyces sp.  
US-11-143-980-49

Query Match 3.6%; Score 126; DB 7; Length 7968;  
Best Local Similarity 23.3%; Pred. No. 3.8;  
Matches 186; Conservative 79; Mismatches 285; Indels 248; Gaps 45;

Qy 15 RRVFRDRRGRVLAELIAAYRDOPVIVGLARGLFPVMEVA-AALHAPLDAFVVRKL 73  
Db 7137 REVVRKAKADGARHLHELTRDI-DLDAFVLFSSAAGTVGNAGGSYAANAVALDGLAMPR- 7194  
Qy 74 GAPG--HDEFVAGLASGRVAVNDVVRGIRITPQQLRDIAERREGLRLRESAYRGER 131  
Db 7195 RABGLVATSVAMGMAESGMAA-----EMARSGMD-PRSAALAAGLV 7236  
Qy 132 PPTDITGTVIVVDGLATGASMFPAVQ-----ALRDAPOAQ-IVIAVP 174  
Db 7237 LAABET--TVVVAIDIMATIPGARFTASRPSLSELSDGSVSTEAADGERPADAFATLE 7294  
Qy 175 AAEPTCREPAGLVDDVVCATMPT-----PLTAGESFMDPRQVTDEEVR 220  
Db 7295 AMAE---RERAAVTLDLVTRTHAAVLGHTASEAIDPAPFOEIG-----FDSLTAVEIRN 7346  
Qy 221 LLATPTGASLRRAA-----STADVLRVAIDA--PGGVPTHEVLAELVGDARIYVLIG 273  
Db 7347 RL---TAATGVRFPAASYIVDYPTPALALAHVCREALGCGRTPAVVPRPVDDDEFIALIG 7403  
Qy 274 ESSHGTHFYOARAAMTQ--WLIBEKGFGAVA--AEADW-----PDA-YRVVRYYR- 319  
Db 7404 MSCR-----FPGGVSSPEDMGLAEGRDANSDFPADGCMULAEIYDDDPHPSSTYRA 7458  
Qy 320 -GLGSDTNA-----DEALSGFER--FPAMMRNTVVRDVF 351

Db 7459 CGFLDDAAADPDGPFCTISPREBALAMPQQRLLLEVAMEAERAMSPATL-KGSRGVFV 7517  
Qy 352 EWLTRNRQRY---ESGALROA-GF-----YGLD-----LY 377  
Db 7518 ---CTNQDVAALASGAPRGAEGYLGTSAAASYASGRLATVFGLEGPAVTVDTACSSSLV 7574  
Qy 378 SLRSTOEVISYLDKVDPRRAAPARARYCF---DHACDDGQAVGFPA--FGAGP 429  
Db 7575 ALHIAAQLRSGECSLALAGATVMATPAAFLERSRRLAALAGRCFAAALDGTGWE 7634  
Qy 430 SCREAAVEQVLDVORN---ALAYAR-----ODGLAEDELFYAQNAQTTRDAEYVRAM 481  
Db 7635 GVGMLTVERLSDDERNHRVLAVVRGSAVNODG--ASNGLTAPNGSQ-----QRVTRQAL 7688  
Qy 482 FSGRVTSMNLRDQHMAQTLSLTLTHLDRHLDAAPPARIYVVAHNSHVGDARATEVWADGOL 541  
Db 7689 ANARLSA-----TDID-----VEAHGT--GTSIGDPLEA----- 7716  
Qy 542 TLGQIVREYXGDESRSIGFSTYGTV-----TAASEMGIAQRKAVRPALHSGVEELF 594  
Db 7717 ---QALATYVG-QGRSQNKPLMGLSVKSNIGHTQAAAGVAGVI--KVMAMMRHGVLPRTL 7770  
Qy 595 HQTADS-----FLVSARLSRDAEAPLDVVRGRALIGVYVLPATERQSHYLVPR 643  
Db 7771 HVDSPPHVDMAARVAVELLVBARBPRTGAP-----RRAGVSSFGVSGTNAHYI---- 7819  
Qy 644 ADQFDAMIHIDQTRALEP 661  
Db 7820 VEGQPVVARPDRESAREP 7837

## RESULT 4

US-11-143-980-48

; Sequence 48, Application US/11143980  
; Publication No. US20050272133A1

; GENERAL INFORMATION:

; APPLICANT: He, Min

; APPLICANT: Hucul, John

; APPLICANT: Haitli, Bradley A.

; APPLICANT: Wegenaar, Melissa M.

; APPLICANT: Graziani, Edmund

; APPLICANT: Summers, Mia

; APPLICANT: Kulowski, Kerry

; APPLICANT: Pong, Kevin

; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Production of a Complex

; TITLE OF INVENTION: Polyketide

; FILE REFERENCE: AM-101426US

; CURRENT APPLICATION NUMBER: US/11/143,980

; CURRENT FILING DATE: 2005-06-03

; PRIOR APPLICATION NUMBER: US 60/664,483

; PRIOR FILING DATE: 2005-03-23

; PRIOR APPLICATION NUMBER: US 60/576,895

; PRIOR FILING DATE: 2004-06-03

; NUMBER OF SEQ ID NOS: 72

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 48

; LENGTH: 7102

; TYPE: PR1

; ORGANISM: Streptomyces sp.

US-11-143-980-48

Query Match 3.5%; Score 122; DB 7; Length 7102;  
Best Local Similarity 22.5%; Pred. No. 6.5;  
Matches 135; Conservative 84; Mismatches 214; Indels 166; Gaps 34;

Qy 82 AVGALASGRVAVNDVVRGLRITPQQLRDIAERBER--ELLRRESAYRGERPPTDITGK 139  
Db 1210 ASGTVAAGGVAPRTEVRVVSSEPDV---VGSAVGHVLEVLBRVQAMVDE--DLAGE 1263  
Qy 140 TVIVV-----DDGLA--TGASMFPAVQALRDAQRAQIYIAPAPAEPTCREPAGLVVD- 169  
Db 1264 RLVVVTRGAVDTGADVADVGAALVAGLVRSAQSENPERLV-----LVDT 1307



QY 190 ---DVGCATMPPLAVGESFMDRQVTDDEVRLLATPTAGPSLR-PAASTAADVLR 245  
| : : : : :  
DB 1308 DDGLGVSLP-GMLAT-----DEEQIVSRGAVVPRIARVAPAGVSGFGS 1355  
QY 246 VAIDAGGVPTHEVLAELV-----GDARIVLI-GESSHGTHEFYQARAAMTOMLI 294  
| : : : : :  
DB 1356 GAVLVYGTG---GVLGGIVSRHLVARRGVSKVLVLSRGAFAEGAALREB-----L 1404  
QY 295 EEKFGAANAADMPDAYRVNRVYRGJEDTNADEALSGFERPPAMWMTVVDPEWML 354  
DB 1405 EAAGAEVVIACDAADREALAGVLSGL---SADFALSG-----VV----- 1441  
QY 355 RTRQRRESGALRQAGFGGLDLYSLHRSIQEVIISLDKVDPRRAARARACDHAACAD 414  
DB 1442 ---HAAGVLDL---GL-LTSLTR-----BEVEPVLRKVDAAANNLHETLTM 1481  
QY 415 DGAAYG-PAAAF-AGPSEREAVEQLVDVQNALVYARODGLAEDELFYAQNQA 469  
DB 1482 DLSAFVLPSSAAGLIGNAGG-----SYAANGFL---DALAAHRRAR 1521  
QY 470 TVRDAEYVY---RAMFSGRVTSWNLRDQH-MAQTIGSLTLTHDRHLDAEPARIYVMAH 523  
DB 1522 GLPAVTSJAMGFWEARSELTOHLSADDLARAHAVPMPTSQALDLPDATTAADEPVVLAAPL 1581  
QY 524 NSHVGDARATEVMD-GOL-TLGOIVRERYGDESRSIG---ESTYGTVTA---SENG 574  
DB 1582 N-----PQAWSDAGHLPVLRDLVRPRIRAAETTGAPESASALGHRLAAVDRBMD 1633  
QY 575 GIAOKKAVRPALPHSGVEELFHQTADSFVLSARLSDAEPDLV-VRLGATGVVLYPAT 632  
DB 1634 QVV-RELVRNHIAVLRHSGESVDTSTRTFOETGFDLSITAVELNRRISAATG-VRLPAT 1690

RESULT 5  
US-10-858-730-89  
; Sequence 89, Application US/10858730  
; Publication No. US20050255568A1  
; GENERAL INFORMATION:  
; APPLICANT: Bailey, Richard B.  
; APPLICANT: Blomquist, Paul  
; APPLICANT: Doten, Reed  
; APPLICANT: Driggers, Edward M.  
; APPLICANT: Madden, Kevin T.  
; APPLICANT: O'Leary, Jessica  
; APPLICANT: O'Toole, George  
; APPLICANT: Trueheart, Joshua  
; APPLICANT: Walbridge, Michael J.  
; APPLICANT: Yorgey, Peter S.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID  
; FILE REFERENCE: 14184-030001  
; CURRENT APPLICATION NUMBER: US/10/858, 730  
; CURRENT FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/475, 000  
; PRIOR FILING DATE: 2003-05-30  
; PRIOR APPLICATION NUMBER: US 60/551, 860  
; PRIOR FILING DATE: 2004-03-10  
; NUMBER OF SEQ ID NOS: 364  
; SOFTWARE: PasteSeq for Windows Version 4.0  
; SEQ ID NO 89  
; LENGTH: 528  
; TYPE: PRT  
; ORGANISM: Mycobacterium leprae  
US-10-858-730-89

Query Match 3.5%; Score 121.5; DB 6; Length 528;  
Best Local Similarity 22.6%; Pred. No. 0.17; Indels 183; Gaps 27;  
Matches 118; Conservative 48; Mismatches 174;

QY 30 ELAAAYRQDPVIVYGLARGGLPVAMEVAALHAFLPAFVYRKLGAGHDEFAVGALASG 89  
| : : : : :  
DB 36 KLAAAV---PADAL-LVRSATTVDAEVLAA--APKLIYAR--AGVGLDNVDVDAATAR 87

QY 90 GRVYN-----DVRVGRITPQOLRDIAREG---RELRRESAYRGRPPTDITGKT 140  
| : : : : :  
DB 88 GVLVNAPTSNIHSAEHALALTLAASRQIAEDADSLRAHITGRSSFSG---TETFGKT 143  
QY 141 VIVVDG---LATGASMPAAVQALRDAPQAQ-----IVIAV 173  
DB 144 VGVVGLGRIGQLVAARIAFGAHVIAVDPVVAPARAQIGIELMSFDLLARADFTSVHL 203  
QY 174 PAAPSTCEBAGLVD-DVVCATMPPTFL-----AVGESFW 208  
DB 204 PKTPET-----AGLIDKEALAKTKPGVIVYNAARGLVDEVALADAVRSGHVAAAGLDVF 258  
QY 209 DFRQVYD---EVRRLAPPTAGPSLRRAASTAAVLRV-----AIDRPGVY 254  
DB 259 ATPRCTDPSLFEISQVVVYTHLGASTAEADDRAGTVASVRLALAGEFVPDVAIVNDGV 318  
QY 255 PTHEVLAELVDGARIVLIGSSHGTHEFYQARAAMTOM-LIEKFGAANAADMPDAY 312  
DB 319 VNEEV-----APWLDVYKGLVVAALSDDELPAUL 348  
QY 313 RVNRVYRG--LGEDTN-ADAEALSGFERPPAMWMTVVDPEWMLTRNQRYSGALRQ 368  
DB 349 SV--HYRGELASBDEILRLSLRGL-----FSTVEDAVTFVNAPALAAERGVSAE 398  
QY 369 --AGFGDLVSLHRSIQEVIISLDKVDPRRAARARACFDHACADGQAYGFAARG 426  
DB 399 ITTSSSPN---HRSVVDV-----RAVASGGSVVNIAGTL- 430  
QY 427 AGPSEREAVEQLVDVQ-NV-----ALAYARODGLAE 458  
DB 431 SGP-----QLVOKIVQVNGNRPDLRAQGMNLVIRYVDPQALGK 469

RESULT 6  
US-11-205-109-15  
; Sequence 15, Application US/11205109  
; Publication No. US20050287641A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Zazopoulos, Emmanuel  
; APPLICANT: Staffa, Alfredo  
; TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS  
; FILE REFERENCE: 3002-2US  
; CURRENT APPLICATION NUMBER: US/11/205, 109  
; CURRENT FILING DATE: 2005-08-17  
; PRIOR APPLICATION NUMBER: US/09/976, 059  
; PRIOR FILING DATE: 2001-10-15  
; PRIOR APPLICATION NUMBER: US 60/239, 924  
; PRIOR FILING DATE: 2000-10-13  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 15  
; LENGTH: 8695  
; TYPE: PRT  
; ORGANISM: Actinoplanes sp.  
US-11-205-109-15

Query Match 3.3%; Score 115.5; DB 7; Length 8695;  
Best Local Similarity 21.7%; Pred. No. 27;  
Matches 193; Conservative 65; Mismatches 294; Indels 339; Gaps 46;

QY 31 LLAAYRQDPVIVYGLARGGLPVAMEVAALHAFLPAFVYRKLGAGHDEFAVGALASG 75  
| : : : : :  
DB 7660 LLDVHRDG-----VGLVQERPEJDGEVVARLRA---VARRIGVSPATVMVAMARVLG 7709  
QY 76 --PGHDEFAVGALASGRVNVNDVVRG--LITTPQOLR-----DIAE--REGRELL 121  
DB 7710 VISGRDVFVFTLL-GRFSTGADRVGPPIVTLVPRAVLGTTGTAAGAAVAEMRRLAELL 7768  
QY 122 RRE-----SAYRGERP-----PTDITGKT----- 140  
DB 7769 EHEHAPLTTAQAASGLSGSLPLFTALFVNRHNTSPGADSPAAAGPTEGIRPVSMERTVY 7828

```

Qy 141 ---YIVVDDGATGASMFPAVQALRDAPQIVAVPAAPSTCREPAGLVD----- 189
Db 7829 PISVAVDDDEGLG-----VAVNAIPVPEAVCELVATATESTLSALELFDGSGPDPAVG 7884
Qy 190 ---DVV-----CATMPTPLAVGSEST-WDFQV-- 213
Db 7885 ELDDLPRGERSRLLEWMDTKARPVVSESVPLAFERVAALAAADAVAVGEGVSWSTREIDR 7944
Qy 214 -TDEEVRLATPTAGPSLRPP--AASTADVLR-VAIDAPGV-----PTHEVL 260
Db 7945 RSDVLAISVA--AGVGLSEPPVVALLERSADVLAFLAVAKAGVFPVVDLSPPQVRID 8001
Qy 261 AELVGDARIVLIGSSSGTHEFYQARAAMTOMLIEKGFGAVAAADWP-----DAYRV 314
Db 8002 A-VIADSRPVLVLD-----VDLPAAEADLPFVPAGAGVVRM 8037
Qy 315 NRYVRG-----IGEDT-----MDEAL--SGFERF----- 337
Db 8038 --YTSGSTGRPKGVVTHQNLVDLATPTCWGSTPRVLPHAPHADASSYEIVWPLNGT 8095
Qy 338 ----PAMWRTVVRDFEWMILRTNORYESGALRQAGFYGLDLSLRSIQEIVSYLDKV 393
Db 8096 VVAVPRSIDATVARDLVRGHELVHTVYAGLRLV-----LD-PSCFAGLTVETLTCGDV 8149
Qy 394 DPRP-----AARAPARYA-----CFDHACADGQAYGFAAFGAGPSCREAVEQ 438
Db 8150 SAEVRRVKEANPGIARVQLYGPTEVTLCAQHLLDQ-----VPIGRPLNTRYVV 8201
Qy 439 LVVDQR-----ALAYARQDGLAE-----DELFFAQQMAQTVR 472
Db 8202 LDDLRFVPTGVVGLVYVAGSLARGYAGMPGLTAERFVADPFSVGGSLYVTGDLVMTD 8261
Qy 473 DAEEYRAMEFGRVTSWNLARDQHAQTLGSLTLHRLDLPARIVVMANSHVGDARA 532
Db 8262 DGVLA-----PAGRAD--QVKRGYRVEGEVETALQHPVDSQVAVVV--REDAPGDKRL 8314
Qy 533 TEVWADQQLTGQIVREYRGDSRSIGFSTYGTV-----TA-----ASEWCG 575
Db 8315 VAVYVGDV-----EAYAQE-RLPGYVPSAFVHLALPLTANQKVDRAALPAPERET 8366
Qy 576 IAQRKAVRPAHGSVEELFHTADSFLVSARLSRDAERPLD-----VRL 620
Db 8367 TTPGKAAPGPGNLLESMSQA-----FAEVLGLDSVGPDDDFPAGGSHLAAVALVQL 8421
Qy 621 ---GRAIGVYVLPATERQSHYL-----HVRPADQPDAMIHIQTRALEPL 662
Db 8422 KARQAVTVQDIMAAPVSELMGSLMSSTR--DSLGLTLPFRKTGELPL 8470

RESULT 7
US-11-055-822-898
; Sequence 898, Application US/11055822
; Publication No. US20050260707A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Krieger, Burkhard
; APPLICANT: Schröder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; FILE REFERENCE: BGI-121PCN
; CURRENT APPLICATION NUMBER: US/11/055,822
; PRIORITY FILING DATE: 2005-02-11
; PRIOR APPLICATION NUMBER: 09/606,740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141,031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/142,101
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: 60/148,613
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/187,970

```

```

; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: DE 19930476.9
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: DE 19931415.2
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931418.7
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931419.5
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931420.9
; PRIOR FILING DATE: 1999-07-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1158
; SEQ ID NO 898
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-11-055-822-898

Query Match      3.2%; Score 113; DB 7; Length 449;
Best Local Similarity 22.9%; Pred. No. 0.58;
Matches 63; Conservative 32; Mismatches 90; Indels 90; Gaps 11;

Qy 15 RRYPRDRREGRYLAELIAYRDQPDVIVGLARGGLPVAVWEVAAALHAFLDAFVVRKLG 74
Db 231 RVNVEARLEIGRKIALEAPAVGD---LVITPESGTPAVVGFPAQASGIFGQGVKN-- 284
Qy 75 APGHDEFAVGALASGGRVVVN--DDVVR--GLRITPOQLRDIARREGRELLRBSAYRGER 131
Db 285 -----HYVGTFIQPSDTLRQLGIRLKNLPRV----- 313
Qy 132 PPTDTGKTVIVVDDGATGASMFPAVQALRDAPQIVAVPAAP-ESTCREPAGLVD 190
Db 314 ---IAGKRLVVVDDSVIRNTQRAVIRMLREAGAAEVHVRIRASPVKWC--FYGI--- 364
Qy 191 VVCATMPTPLAVGSESWDRQVYTDDEVRLTLPTTAGPSLRPPASTADVLRVAIDA 250
Db 365 -----DPAITGELIAN-AVTSIDNEAEVVEAVRSA-----IGADTLGYSID 405
Qy 251 -----PGVPTHEVLAELV 264
Db 406 MVAITEQANELCTIACFDGKTPWGLPQGNANADLV 440

RESULT 8
US-11-075-185-7
; Sequence 7, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMERUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 7
; LENGTH: 3507
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-11-075-185-7

Query Match      3.2%; Score 111; DB 7; Length 3507;
Best Local Similarity 21.5%; Pred. No. 16;
Matches 177; Conservative 79; Mismatches 244; Indels 322; Gaps 48;

Qy 37 DQPDVIVGLARGGLPVAVWEVAAALHAFLDAFVVRKLGAGHDEFAVGALASGGRVVND 96

```



APPLICANT: He, Min  
APPLICANT: Hucul, John  
APPLICANT: Haultli, Bradley A.  
APPLICANT: Wagenaar, Melissa M.  
APPLICANT: Graziani, Edmund  
APPLICANT: Summers, Mia  
APPLICANT: Kuliowski, Kerry  
APPLICANT: Pong, Kevin  
TITLE OF INVENTION: Biosynthetic Gene Cluster for the Production of a Complex  
FILE REFERENCE: AM-101426US  
CURRENT FILING DATE: 2005-06-03  
PRIOR FILING DATE: 2005-03-23  
PRIOR APPLICATION NUMBER: US 60/564,483  
PRIOR FILING DATE: 2004-06-03  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 47  
LENGTH: 5712  
TYPE: PRF  
ORGANISM: Streptomyces sp.  
US-11-143-980-47

Query Match 3.1%; Score 109.5; DB 7; Length 5712;  
Best Local Similarity 20.1%; Pred. No. 42;  
Matches 157; Conservative 86; Mismatches 265; Indels 275; Gaps 42;

19 RDRBRAGVLAELAAAYDQDPVYVGLAR--GGLPYAME--VAALALPLDAFVVRK- 72  
864 RDCLTGPGTLVGTLRDRPEFOALVTALAEVYSGVEVAMSPVSGGRRIPLPTAFQRC 923  
73 ---LGAPGHDFAVGALASGRVVDVVGRLRTTPQQLDIAREGRELRESAARG 129  
924 RYMFSAPEPE---SG-----TTPGHGVT-----SGRE--KDTGICSG 955  
130 ERPEPTDITGKTVIVDDGATGAMFAAVQALRDAQPAQIVYAPAA---PESTCREFA 185  
956 DEAF-----DTG-PGGGELTGMVRA-----HAAVVLGYASATAIGAHTFFQKG 998  
186 GLVDVVCATMPPTPLAVGESFMDPROVTDEVRRLATPTAGP-----SLRPPAAS 238  
999 -----FDSITAVELCERLCAATLPLPGLTLPDYTPAA-L 1033  
239 AADVLARV-----AIDARGVTHEVLAELVGDARIYVIGSSHCTHEFYQARAAMQWL 293  
1034 AEHLHRRHLHGRTDEQAAPATVPTPD-----GGDPVIVVGCCR-----FGRASPE-- 1080  
294 IEKGFCAVAABADWPDAYRVNRVYVRLGSDTNADALSGFERFPAA---MMRTVVRD 349  
1081 -----DLMRI-----VADGEDAISGFPDRGMWLAGLYHPDRPH 1115  
350 FVEMLRFTNQYESGALRQAGFYGLDLYSLHRSIOEVIYSLDKVDPRAABARARYACFD 409  
1116 GTSYARDCGFLYDAEF--DAGFFGI-----SPREA---EAMPDQORLLLETSMWALE 1163  
410 HACADDCGAYFAAAGFAGPCEREAIVQYDVQRMALAYARODGLAEDLEFPAQONQ 469  
1164 RA-----GIPAEHIKSSGTG-----VFIGASSGVYADAGEAEG--YOUTGTAA 1206  
470 TVRDAEYVYRAMFGS-----RVTSMNLRDQMA-QTIG-----SILLTHLDHRLDAPARI 518  
1207 SVASGRVSYTIGLEGPATVYVJTACSSSLVALHIAVQSLRAGECSIALAGCVTVAATPAMF 1266  
519 VVMAHNSHVG--DAPATEV-----WADGQLTLGQIVREERYGDESR-----S 557  
1267 VEFSSRQGLAMDGRCFAAADAADGTGMAEG--VGLVVERLSDAERNGHRLAVVRGSA 1323  
558 IGFSTYCTGTVAASWKGIAQRKAVRPL-----HGSVELEFHC-TADSF 601  
1324 VNODGASNGLTAPN---GPSQORVIRQALASAGLVASDVDAVEAHGTGTTGIDPLEAQL 1380

602 LVSAKLRDAEAP-----DVARLGAIGVYVLPATEROSHLYHR 642  
1381 LATYGGRDADRPIMLGVSXNIGHTOAAAGVAGVITVMAVRGVLPRT-----LHVD 1434  
643 PADQDAMIHD-----QTRALEPLEVTSRMWINGEN-----PET-Y 677  
1435 -----EPSTHVDMSGRVELLTGTTPWPTTGLRRAGVSSFGVGNAAHYLEQVETAR 1489  
678 PTG 680  
1490 PTG 1492

RESULT 11  
US-11-075-185-8  
Sequence 8, Application US/11075185  
Publication No. US20050266434A1  
GENERAL INFORMATION:  
APPLICANT: REEVES, CHRISTOPHER D  
APPLICANT: JULIEN, BRYAN  
TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS  
FILE REFERENCE: 010099.03  
CURRENT FILING DATE: 2005-03-07  
PRIOR FILING DATE: 2004-03-07  
PRIOR APPLICATION NUMBER: US 60/551,103  
PRIOR FILING DATE: 2004-03-08  
PRIOR APPLICATION NUMBER: US 60/566,290  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 8  
LENGTH: 2197  
TYPE: PRF  
ORGANISM: Sorangium cellulosum  
US-11-075-185-8

Query Match 3.1%; Score 109; DB 7; Length 2197;  
Best Local Similarity 21.9%; Pred. No. 12;  
Matches 95; Conservative 47; Mismatches 159; Indels 132; Gaps 21;

2 LMTAADVTRSPRRVRD--RREAGVLAELAAAYDQDPVYVGLARGLPYAMEVAAA 60  
1630 LETVAACARCAIGVSVYSGELHADTGELVS--ARLRKSPVVYLD-----EAIAG 1676  
61 LHMPLDAFVVRKXGAPG---HDEFAV---GALASGRVVDVVGRLI-----TPQ 107  
1677 FELPFGVCAIRRFHESGVMTRQPEEFARVGVSGWAGPALTVBENILRRFRAAVVTTNDTIA 1736  
108 QLRDIA--EREGBELR-----RESAYRGERPPTDITGKTVIVVD 145  
1737 NLRALAVDQRTKEAHNSYVNPVLESIDAFGLAGRHRHADRGEYLERD--DSSARVIN 1795  
146 DGLATGSMFAVQALRDAQPAQIVYAPPAPESTCEFPAGLVDDVVCATMPPTPLAVGB 205  
1796 LYVITSAS-----FRGHTGSEIAGSVLGTHTDITRDYMDLEBRIPRET----- 1838  
206 SFMDPROVTDEVRRLATPTAGPSLRRAASTAADVLARVALDAPGCVTHEVLAELV 265  
1839 ---DFGRV-----FPAAGP-----ATAVETAVNKLGL-----LAARKG 1867  
266 DARVILGSSSHGTHEFYQARAAMTQWLIEKGFCAVAABADWPDAYRVNRVYVRLGSDT 325  
1868 -SALLVLGS-----PFTFGLGLVSHAEFGSPLEALVESCPMSKVIAVDPFGGAAAE 1921  
326 NA-----DEALSGFERPAMRMRTVVDVFEWMLRTNRQYESGALRQAGFYGLDLYSLR 381  
1922 EAKLTSDDV--GF-----VMLETLQSD--W-----GGLRSVPDAVLEVIDRHR 1960  
382 SIQEVISYLDKVD 394  
1961 ---ERSGYLVGD 1970



NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 20147  
; LENGTH: 876  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(876)  
; OTHER INFORMATION: Ceres Seq. ID no. 12377951  
US-11-096-568A-20147

Query Match 3.1%; Score 107.5; DB 7; Length 876;  
Best Local Similarity 21.2%; Pred. No. 4;  
Matches 125; Conservative 70; Mismatches 207; Indels 187; Gaps 31;

QY 142 IVVDDGLATGASMFVAOALR-----DAOPAOIVIAVPAPESTC-----REFAGLVDDV 191  
DB 48 MVLSDEHTTMSMLTAENAVRDGSIKGSIVHL-----ECTGSTISRIIIVIKLDV 103  
QY 192 V---CATM--PTPLAVGESFWDPROVTEDEVRLLA-----TTPAGPS-LRRPAS 237  
DB 104 LQSCAIGNRPY-----EMRQNEGCTNIPANAQANTGTVSSGFLGSPAS 155  
QY 238 TAAVLRVAIDAQGG-----VPTHEVLAVGD-AAIVLIG-ESSHGTETFOQAPAM 289  
DB 156 RPAOVANNVPYSGSYGHTGVPEIAPAVESVPVAVSGVTYGTSAHT-----MNAGM 210  
QY 290 TQMLIEKGFVAVAEADWPDAVRYNRVVRGLGE--DTNADALSGEPFPAMMMRNTVV 347  
DB 211 TQSNLQQRSLNSHPNQ-----RFAVPSMAGSGAPAGNYGPPQAPPGHMRTPV 265  
QY 348 RDPFWEMLTRNQRYESGALRQAGFY-----GLDLYSLHRSIOEIVISYLDKV 393  
DB 266 -----SKNDARLVPLAQLNPYLDKWTIKVRYTAKTDLRFNNMAGAKVPSF-DLL 316  
QY 394 DPRAPARARAYACD-----RKEIRATCFNTQADQFNLIEVDKYLLISKSLKPAOKKFNLSNHEYSILDSR 409  
DB 317 DEQ---RKEIRATCFNTQADQFNLIEVDKYLLISKSLKPAOKKFNLSNHEYSILDSR 373  
QY 410 ---HACADG-----QAYGFAAFGAGPSCEBEAVEOLVDYQRN--ALAVARODGLLAE 458  
DB 374 TSIEVCADDDSNIPROQYNF-RKISEINIEKDLIDLIGTVTSVGSVTTIRKQGV--- 429  
QY 459 DELFYAQOACTVDAEYVYRAMPFSGR---VTSWNLBRDQNAQTLGSL-LTHLDRHLDPAR 514  
DB 430 ---ETQRRTLELD-----MSGRSVQLTWGKLCVAREGNOQLSDGSLMPVLAALK 477  
QY 515 PARIVVAHNSHVGDARATEVWADGQLTLGQIVREERYGDESRSIGFTYGTTPAS--- 571  
DB 478 GARVTVDSGRS-VESAGSTOLKIDPEIPEASLRRWYA-----TGCKTAAACVSL 525  
QY 572 ---EMGGIAORKA---VRPALHGSVEE-----LPHQTADSLVSA 605  
DB 526 SVSMGTCVRSIAQIDENLGOLEKDPITVKAALSHLADNFCYPA 574

## RESULT 15

US-11-075-185-6  
; Sequence 6, Application US/11075185  
; Publication No. US20050266434A1  
; GENERAL INFORMATION:  
; APPLICANT: REEVES, CHRISTOPHER D  
; APPLICANT: JULIEN, BRYAN  
; APPLICANT: REID, RALPH  
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS  
; FILE REFERENCE: 010099.03  
; CURRENT APPLICATION NUMBER: US/11/075,185  
; PRIOR FILING DATE: 2005-03-07  
; PRIOR APPLICATION NUMBER: US 60/551,103  
; PRIOR FILING DATE: 2004-03-08  
; PRIOR APPLICATION NUMBER: US 60/568,290  
; PRIOR FILING DATE: 2004-05-04  
; NUMBER OF SEQ ID NOS: 61

SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 6  
; LENGTH: 1892  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-11-075-185-6

Query Match 3.1%; Score 107.5; DB 7; Length 1892;  
Best Local Similarity 24.0%; Pred. No. 12;  
Matches 131; Conservative 42; Mismatches 187; Indels 185; Gaps 27;

QY 2 LMTAAAVTRRSPPRVRD--RRAGRVLAELAAVDDQDVIVLGLARGGLPTVAEVA 59  
DB 1286 LLEALAD-GAAPPRLVVDLTYAGDAGVVAVHAAVRG---ALALVG---WLAAP 1334  
QY 60 ALHAPLDAFVVRKUGARSHS-----FAVGALASGRVVVNDVVRGRLITPOQ--- 109  
DB 1335 QLAATELVVTRCAVAVGPEGVDAIGPAVWGLLRTRTAEPYRAVRVLDVGRNPLDGA 1394  
QY 110 -----RDIAREGRELRRSAYRGERPPTD-----ITGKT----- 140  
DB 1395 LRRALAAGTPELVSRSGBARAPRLREVRGSEBPAAAPATRLDPGTALITGTGELGRH 1454  
QY 141 -----VIYDDGLATGASM--FAVQALRDAQOQIVT 171  
DB 1455 VAKHLVTAHGVRLVLTSSRGMDAPDAALVDELRAGAATVDVAACDADAAALAAVVE 1514  
QY 172 AVPA--PESTCRPEAGLVDDVVCATMPTPLAVGESFWDPROVTEDEVRLLATPTAGPS 230  
DB 1515 AIFAPRLTAVHTAGVLDSDSVT-----KLSAEQLAKVL----- 1549  
QY 231 LRRPAASTADVLRRAVIDAPGVPTHEVLAEVGDARIVLIGSSHGTETFOQAPAMT 290  
DB 1550 -----RPKVD--GAFTLHE-LTKAPLAAPVLF--SSAAGT----- 1580  
QY 291 QMLIEKGFVAVAEADWPDAVRYNRVVRGLGEBTNADEALSGEPFPAM-MRNTVV-- 347  
DB 1581 ---LGSFGQANYAAANTFLDALASHLRARGV-----PAMS-----LAWGFMAQTGIGM 1625  
QY 348 ---RDPFWEMLR-----TRNQRYESGALRQAGF-----XGLDLYSLHRSIOEIVISY 389  
DB 1526 TAILGAADIDARKHGVSPVAVAGRLRLDRALAAQBAATVPLALDSSIQRA-----GSN 1681  
QY 390 LDKVDP--RAAARARAYACFDHACDDGQAYGFAAFGAG---PSCEREAVEQVLVDQ 443  
DB 1682 AGVPVPLRLGLVRAPARRTAASAAANGNGTG-AAALRLRLSPLPGAERQKV--LLDLV 1738  
QY 444 RNALA 448  
DB 1739 RTBIA 1743

Search completed: March 23, 2006, 06:29:34  
Job time : 19.3423 secs

***This Page Blank (uspto)***

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using SW model

Run on: March 23, 2006, 04:56:29 ; Search time 38.2848 Seconds  
(without alignments)  
1652.628 Million cell updates/sec

Title: US-10-617-038-24  
Perfect score: 736  
Sequence: 1 MATTLPIVQRHPSLHPEFSE.....SVAVSEKPKTEKHQIRSTN 144

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_21:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*
- 9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	736	100.0	144	2	AAW94900 Alpha-cry
2	736	100.0	144	5	AAW50751 Mycobacte
3	736	100.0	144	8	ADI37303 M. tuberc
4	736	100.0	330	5	AAU74589 Antigenic
5	736	100.0	331	2	AAV32060 Mycobacte
6	736	100.0	331	7	ADF69755 Fusion pr
7	736	100.0	368	2	AAV32069 Mycobacte
8	736	100.0	368	5	AAU74598 Antigenic
9	736	100.0	368	7	ADF69775 Fusion pr
10	736	100.0	433	5	AAV32065 Mycobacte
11	736	100.0	433	7	AAU74594 Antigenic
12	736	100.0	433	7	ADF69765 Fusion pr
13	736	100.0	856	2	AAV32064 Mycobacte
14	736	100.0	856	7	ADP69763 Fusion pr
15	736	100.0	859	5	AAU74593 Antigenic
16	736	100.0	875	7	ADA26365 Mycobacte
17	718	97.6	143	9	ADZ47757 Mycobacte
18	356	48.4	131	2	AAK12489 Peptide d
19	345	33.3	48	2	AAK85672 12 KD M.t
20	245	33.3	48	2	AAK92889 Mycobacte
21	245	33.3	48	2	AAW18183 N-termi
22	245	33.3	48	2	AAW75570 M. tuberc
23	245	33.3	48	6	ABG74430 M. tuberc
24	245	33.3	48	6	ABU56334 M. tuberc

25	245	33.3	48	7	AAE39277 M. tuberc
26	245	33.3	48	7	ADF45119 M. tuberc
27	245	33.3	48	8	ADO36804 12KD majo
28	245	33.3	48	8	ADU64179 Mycobacte
29	164.5	22.4	159	6	ABP57444 Mycobacte
30	156	21.2	146	5	ABG77903 High leve
31	150.5	20.4	188	8	ADQ78294 S. aurant
32	148	20.1	164	9	AEK39411 L. pneumo
33	148	20.1	168	9	AEK35982 L. pneumo
34	146.5	19.9	151	9	ABM94472 M. xanthu
35	143	19.4	158	4	AAU63279 Propionib
36	143	19.4	158	6	ABM59798 Propionib
37	142	19.3	151	8	ADT60218 Plant pol
38	141.5	19.2	154	8	ADM48277 Polypept
39	134	18.2	170	8	ADX74462 Plant ful
40	133.5	18.1	142	7	ADC97715 Streptoco
41	133.5	18.1	154	7	ADC97714 Streptoco
42	133	18.1	151	7	ADP75373 Wheat gHS
43	133	18.1	151	7	ADL18503 Wheat pro
44	133	18.1	193	8	ADY10853 Plant ful
45	132	17.9	139	7	ADC94697 E. faeciu

#### ALIGNMENTS

RESULT 1  
AAW94900 standard; protein: 144 AA.

AAW94900;  
11-MAY-1999 (first entry)

Alpha-crystallin heat shock protein.

Tuberculosis; Mycobacterium; alpha-crystallin heat shock protein;  
acr gene; attenuated; vaccine; antimycobacterial; pathogenic.

Mycobacterium sp.  
W09902670-A1.  
21-JAN-1999.

09-JUL-1998; 98WC-US014227.  
10-JUL-1997; 97US-0052199P.

(USSH ) US DEPT HEALTH & HUMAN SERVICES.  
Barry CE, Yuan Y, Crane D;  
WPI; 1999-120871/10.  
N-PSDB; AAX17761.

Novel attenuated strains of Mycobacterium tuberculosis and M. bovis -  
useful in the production of a vaccine against tuberculosis, achieved by  
reducing or eliminating expression of the alpha-crystallin heat shock  
protein gene.

Example; Page 25; 33pp; English.

The invention relates to a vaccine for protection against tuberculosis in  
which Mycobacterium sp. are attenuated by having the expression of the  
alpha-crystallin heat shock protein gene (acr gene) reduced by at least  
75 percent. The attenuated strains can be used as models for screening  
for novel antimycobacterial agents. Also, the mixture of Mycobacterium  
sp. cellular debris can be used for detection of exposure to M. bovis and  
M. tuberculosis, used for commercial adjuvants, used as diagnostics  
including use as reagents for the development of monoclonal antibodies to  
recognize wild type mycobacteria in patient samples. The vaccine is based  
on M. tuberculosis and is obtained without the multitude of passages so





CC The present sequence represents a low oxygen induced antigen, which is  
 CC used in the exemplification of the present invention.

XX Sequence 144 AA;

Query Match 100.0%; Score 736; DB 8; Length 144;

Best Local Similarity 100.0%; Pred. No. 8.1e-74; Mismatches 0; Indels 0; Gaps 0;

Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATTLPVORHRSLSLPFPESELFAAPSPAGLRPTFDTRLMLBDEMKGRYEVRAELPGV 60  
 DB 1 MATTLPVORHRSLSLPFPESELFAAPSPAGLRPTFDTRLMLBDEMKGRYEVRAELPGV 60  
 QY 61 DDDKVDIMVRDGLTTIKARTEQKDFGRSEFAYGSEFVRTVSLPVGADEDDIKATYDYG 120  
 DB 61 DDDKVDIMVRDGLTTIKARTEQKDFGRSEFAYGSEFVRTVSLPVGADEDDIKATYDYG 120  
 QY 121 ILTVSVAVSEKPTKEKHQIRSTN 144  
 DB 121 ILTVSVAVSEKPTKEKHQIRSTN 144

#### RESULT 4

AAU74589 ID AAU74589 standard; protein; 330 AA.

XX AAU74589;

DT 29-AUG-2003 (revised)  
 DT 08-MAY-2002 (first entry)

DE Antigenic fusion protein Erd14-DPV-MTI.

KW Fusion protein; tuberculosis; Mycobacterium tuberculosis;  
 KW tuberculostatic; immunogen; vaccine; Erd14-DPV-MTI; Erd14; DPV; MTI.

OS Mycobacterium tuberculosis.

OS Chimeric.

PN US2002009459-A1.

PD 24-JAN-2002.

PF 07-APR-1999; 99US-00287849.

XX 13-MAR-1997; 97US-00818112.

PR 01-OCT-1997; 97US-00942578.

PR 18-FEB-1998; 98US-00025197.

PR 07-APR-1998; 98US-00056556.

PR 30-DEC-1998; 98US-00223040.

XX (REED/) REED S G.

PA (SKEIT/) SKEIT Y A.

PA (DILL/) DILLON D C.

PA (ALDE/) ALDERSON M.

PA (CAMP/) CAMPOS-NETO A.

PI Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;

XX WPI; 2002-171134/22.

DR N-PSDB; ABA14129.

XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for

PT diagnosing, treating or preventing M. tuberculosis infection,

PT particularly as vaccine for treating or preventing tuberculosis.

XX Claim 1; Fig 2; 62pp; English.

CC protection in animals against the development of tuberculosis. The  
 CC protein coding sequence may be used to encode a protein product for use  
 CC as an immunogen to induce and/or enhance an immune response to M.  
 CC tuberculosis. This sequence represents an M. tuberculosis fusion protein  
 CC of the invention. (Updated on 29-AUG-2003 to standardise OS field)

XX Sequence 330 AA;

Query Match 100.0%; Score 736; DB 5; Length 330;

Best Local Similarity 100.0%; Pred. No. 2.7e-73; Mismatches 0; Indels 0; Gaps 0;

Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATTLPVORHRSLSLPFPESELFAAPSPAGLRPTFDTRLMLBDEMKGRYEVRAELPGV 60  
 DB 8 MATTLPVORHRSLSLPFPESELFAAPSPAGLRPTFDTRLMLBDEMKGRYEVRAELPGV 67  
 QY 61 DDDKVDIMVRDGLTTIKARTEQKDFGRSEFAYGSEFVRTVSLPVGADEDDIKATYDYG 120  
 DB 61 DDDKVDIMVRDGLTTIKARTEQKDFGRSEFAYGSEFVRTVSLPVGADEDDIKATYDYG 127  
 QY 121 ILTVSVAVSEKPTKEKHQIRSTN 144  
 DB 128 ILTVSVAVSEKPTKEKHQIRSTN 151

#### RESULT 5

AAV32060 ID AAV32060 standard; protein; 331 AA.

XX AAV32060;

DT 17-JAN-2000 (first entry)

DE Mycobacterium tuberculosis antigen fusion protein Mtb39A.

KW Tuberculosis; antigen; fusion protein; Mtb39A; ERD14; DPV; MTI;

KW diagnosis; therapy; vaccine; immunogen.

OS Mycobacterium tuberculosis.

PN W09951748-A2.

PD 14-OCT-1999.

PF 07-APR-1999; 99WO-US007717.

PR 07-APR-1998; 98US-00056556.

PR 30-DEC-1998; 98US-00223040.

XX (CORI-) CORIXA CORP.

PA Skeiky YAW, Alderson M, Campos-Neto A;

PI WPI; 1999-601610/51.

XX N-PSDB; AAZ20195.

XX New fusion proteins useful for diagnosis, prevention and treatment of

PT tuberculosis.

XX Claim 1; Fig 2; 83pp; English.

XX This sequence represents a recombinant Mycobacterium tuberculosis tr1-

CC antigen fusion protein, termed Mtb39A, composed of the antigens ERD14,

CC DPV and MTI. The fusion protein is expressed in host cells using a vector  
CC carrying a polynucleotide (see AA20195) comprising the 3 coding  
CC sequences for the antigens. The invention provides fusion proteins (see  
CC AA32059-71) containing at least 2 M. tuberculosis antigens. The new  
CC fusion proteins and polynucleotides encoding them are useful as vaccines  
CC for preventing tuberculosis (claimed), for diagnosis (via in vitro assays  
CC or intradermal skin tests for detection of anti-M. tuberculosis  
CC antibodies), monitoring of disease progression, and treatment of  
CC tuberculosis. They are more effective immunogens than mixtures of the  
CC individual protein components

CC  
XX  
SQ Sequence 331 AA;

Query Match 100.0%; Score 736; DB 2; Length 331;  
Best Local Similarity 100.0%; Pred. No. 2,7e-73;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATTLPVQRHPRSLFPFSESLFAAPSPAGLRPTFOTRLRLDEMKRGYEVRAELPGV 60  
DB 8 MATTLPVQRHPRSLFPFSESLFAAPSPAGLRPTFOTRLRLDEMKRGYEVRAELPGV 67

QY 61 DPKDQVIMVWDGQLTIKAERTEQKDFDGRSEFAYGSFVRTVSLPVGADDDIKATYDKG 120  
DB 68 DPKDQVIMVWDGQLTIKAERTEQKDFDGRSEFAYGSFVRTVSLPVGADDDIKATYDKG 127

QY 121 ILTVSAVSESGKPTKHIQIRSTN 144  
DB 128 ILTVSAVSESGKPTKHIQIRSTN 151

RESULT 6  
ADP69755  
ID ADP69755 standard; protein; 331 AA.  
AC ADF69755;  
AC ADF69755;  
DT 12-FEB-2004 (first entry)  
XX  
XX Fusion protein #2 of M. tuberculosis antigen.  
XX  
XX Fusion protein; Mycobacterium tuberculosis antigen;  
XX tuberculosis infection; immune response; tuberculostatic; mutant; munein.  
XX  
XX Synthetic.  
XX Mycobacterium tuberculosis.  
XX  
XX US2003147911-A1.  
XX  
XX PD 07-AUG-2003.  
XX  
XX PF 05-FEB-2003; 2003US-00359460.  
XX  
XX PR 13-MAR-1997; 97US-00818112.  
XX PR 01-OCT-1997; 97US-00942578.  
XX PR 18-FEB-1998; 98US-00025197.  
XX PR 07-APR-1998; 98US-00056556.  
XX PR 30-DEC-1998; 98US-00223040.  
XX PR 07-APR-1999; 99US-00287849.  
XX  
XX PA (CORI-) CORIXA CORP.  
XX  
XX PI Reed SG, Skeiky VA, Dillon DC, Alderson M, Campos-Neto A;  
XX DR N-PSDB; ADF69754.  
XX  
XX MPI; 2003-897524/82.  
XX  
XX PT New fusion proteins of Mycobacterium tuberculosis for diagnosing,  
XX preventing or treating tuberculosis infection or in enhancing immune  
XX responses in M. tuberculosis.  
XX  
XX PS Claim 1; SEQ ID NO 4; 135pp; English.  
XX  
XX The present invention relates to fusion proteins of Mycobacterium

CC tuberculosis antigens, and the polynucleotide sequences encoding them.  
CC The sequences of the invention are useful in a method for preventing  
CC tuberculosis by administering to a subject an amount of the fusion  
CC protein or the polynucleotide that encodes the fusion protein. Also  
CC disclosed is a pharmaceutical composition comprising the fusion protein  
CC or the polynucleotide sequence encoding it. The fusion protein induces an  
CC immune response to M. tuberculosis and can be used in the diagnosis,  
CC prevention, and treatment of tuberculosis infection. The present sequence  
CC represents a fusion protein of a M. tuberculosis antigen.

CC  
XX  
SQ Sequence 331 AA;

Query Match 100.0%; Score 736; DB 7; Length 331;  
Best Local Similarity 100.0%; Pred. No. 2,7e-73;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATTLPVQRHPRSLFPFSESLFAAPSPAGLRPTFOTRLRLDEMKRGYEVRAELPGV 60  
DB 8 MATTLPVQRHPRSLFPFSESLFAAPSPAGLRPTFOTRLRLDEMKRGYEVRAELPGV 67

QY 61 DPKDQVIMVWDGQLTIKAERTEQKDFDGRSEFAYGSFVRTVSLPVGADDDIKATYDKG 120  
DB 68 DPKDQVIMVWDGQLTIKAERTEQKDFDGRSEFAYGSFVRTVSLPVGADDDIKATYDKG 127

QY 121 ILTVSAVSESGKPTKHIQIRSTN 144  
DB 128 ILTVSAVSESGKPTKHIQIRSTN 151

RESULT 7  
AAV32069  
ID AAV32069 standard; protein; 368 AA.  
XX  
XX AAV32069;  
XX  
XX DT 17-JAN-2000 (first entry)  
XX  
XX DE Mycobacterium tuberculosis antigen fusion protein Mtb36f.  
XX  
XX KM Tuberculosis; antigen; fusion protein; Mtb36f; ERD14; DPV; MTI;  
XX diagnosis; therapy; vaccine; immunogen.  
XX  
XX OS Mycobacterium tuberculosis.  
XX  
XX FH Key  
XX FH Peptide 1..8  
XX FT /note= "Mec/Hls tag"  
XX FT 9..151  
XX FT Protein /note= "ERD14"  
XX FT 154..235  
XX FT /note= "DPV"  
XX FT 238..368  
XX FT Protein /note= "MTI"  
XX  
XX PN MO9951748-A2.  
XX  
XX PD 14-OCT-1999.  
XX  
XX PF 07-APR-1999; 99MO-US007717.  
XX  
XX PR 07-APR-1998; 98US-00056556.  
XX PR 30-DEC-1998; 98US-00223040.  
XX  
XX PA (CORI-) CORIXA CORP.  
XX  
XX PI Skeiky VAM, Alderson M, Campos-Neto A;  
XX DR MPI; 1999-601610/51.  
XX DR N-PSDB; AAZ20204.  
XX  
XX PT New fusion proteins useful for diagnosis, prevention and treatment of  
XX tuberculosis.

PS Claim 1; Fig 11A-B; 83pp; English.

XX This sequence represents a recombinant Mycobacterium tuberculosis tri-  
CC antigen fusion protein, termed Mtb36f, composed of the antigens ERD14,  
CC DPV and MTI. The fusion protein is expressed in host cells using a vector  
CC carrying a polynucleotide (see AA20204) comprising the coding sequences  
CC for the 3 antigens. The invention provides fusion proteins (see AA132059-  
CC 71) containing at least 2 M. tuberculosis antigens. The new fusion  
CC proteins and polynucleotides encoding them are useful as vaccines for  
CC preventing tuberculosis (claimed), for diagnosis (via in vitro assays or  
CC intradermal skin tests for detection of anti-M. tuberculosis antibodies),  
CC monitoring of disease progression, and treatment of tuberculosis. They  
CC are more effective immunogens than mixtures of the individual protein  
CC components

XX Sequence 368 AA;

Query Match 100.0%; Score 736; DB 2; Length 368;  
Best Local Similarity 100.0%; Pred. No. 3.1e-73;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATTLPVQRHPRSLPEPSELFAAPSPAGLRPTFTRLMLBDEMKGREYVRAELPGV 60  
DB 9 MATTLPVQRHPRSLPEPSELFAAPSPAGLRPTFTRLMLBDEMKGREYVRAELPGV 68  
QY 61 DPDKVDIMVWDGQLTKAERTEQKDFGRSEFAYGSFVRTVSLPVGADEDDIKATYDKG 120  
DB 69 DPDKVDIMVWDGQLTKAERTEQKDFGRSEFAYGSFVRTVSLPVGADEDDIKATYDKG 128  
QY 121 ILTVSAVSESGKPTKEKHIOIRSTN 144  
DB 129 ILTVSAVSESGKPTKEKHIOIRSTN 152

RESULT 8  
AAU74598  
ID AAU74598 standard; protein; 368 AA.

XX AAU74598;

DT 29-AUG-2003 (revised)  
DT 08-MAY-2002 (first entry)

XX Antigenic fusion protein Erd14-DPV-MTI (Mtb36f).

XX Fusion protein; tuberculosis; Mycobacterium tuberculosis;  
KM tuberculostatic; immunogen; vaccine; Erd14-DPV-MTI; Mtb36f.

XX Mycobacterium tuberculosis.  
OS Chimeric.

XX US2002009459-A1.

XX 24-JAN-2002.

XX 07-APR-1999; 99US-00287849.

XX 13-MAR-1997; 97US-00818112.

XX 01-OCT-1997; 97US-00942578.

XX 18-FEB-1998; 98US-00025197.

XX 07-APR-1998; 98US-00056556.

XX 30-DEC-1998; 98US-00223040.

XX (REED/) REED S G.  
PA (SKEI/) SKEIKY Y A.  
PA (DILL/) DILLON D C.  
PA (ALDE/) ALDERSON M.  
PA (CAMP/) CAMPOS-NETO A.

PI Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;

XX WPI; 2002-171134/22.

DR N-PSDB; ABK14138.

XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for  
PT diagnosing, treating or preventing M. tuberculosis infection,  
PT particularly as vaccine for treating or preventing tuberculosis.

PS Claim 1; Fig 11; 62pp; English.

XX The invention relates to a purified polypeptide which induces an immune  
CC response of Mycobacterium tuberculosis. Polypeptides of the invention are  
CC useful for diagnosing, treating or preventing M. tuberculosis infection,  
CC particularly tuberculosis infection. In particular, the polypeptides are  
CC useful as a vaccine formulation with an adjuvant to afford long-term  
CC protection in animals against the development of tuberculosis. The  
CC protein coding sequence may be used to encode a protein product for use  
CC as an immunogen to induce and/or enhance an immune response to M.  
CC tuberculosis. This sequence represents an M. tuberculosis fusion protein  
CC of the invention. (Updated on 29-AUG-2003 to standardise OS field)

XX Sequence 368 AA;

Query Match 100.0%; Score 736; DB 5; Length 368;  
Best Local Similarity 100.0%; Pred. No. 3.1e-73;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATTLPVQRHPRSLPEPSELFAAPSPAGLRPTFTRLMLBDEMKGREYVRAELPGV 60  
DB 9 MATTLPVQRHPRSLPEPSELFAAPSPAGLRPTFTRLMLBDEMKGREYVRAELPGV 68  
QY 61 DPDKVDIMVWDGQLTKAERTEQKDFGRSEFAYGSFVRTVSLPVGADEDDIKATYDKG 120  
DB 69 DPDKVDIMVWDGQLTKAERTEQKDFGRSEFAYGSFVRTVSLPVGADEDDIKATYDKG 128  
QY 121 ILTVSAVSESGKPTKEKHIOIRSTN 144  
DB 129 ILTVSAVSESGKPTKEKHIOIRSTN 152

RESULT 9  
ADF69775  
ID ADF69775 standard; protein; 368 AA.

XX ADF69775;

DT 12-FEB-2004 (first entry)

XX Fusion protein #11 of M. tuberculosis antigen.

XX Fusion protein; Mycobacterium tuberculosis antigen;  
KM tuberculosis infection; immune response; tuberculostatic; mutant; mutein.

XX Synthetic.

XX Mycobacterium tuberculosis.

XX US2003147911-A1.

XX 07-AUG-2003.

XX 05-FEB-2003; 2003US-00359460.

XX 13-MAR-1997; 97US-00818112.

XX 01-OCT-1997; 97US-00942578.

XX 18-FEB-1998; 98US-00025197.

XX 07-APR-1998; 98US-00056556.

XX 30-DEC-1998; 98US-00223040.

XX 07-APR-1999; 99US-00287849.

XX (CORI-) CORIXA CORP.

PI Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;

XX WPI; 2003-897524/82.

DR N-PSDB; ADF69774.

PT New fusion proteins of Mycobacterium tuberculosis for diagnosing,  
 PT preventing or treating tuberculosis infection or in enhancing immune  
 PT responses in M. tuberculosis.  
 XX  
 PS Claim 1; SEQ ID NO 24; 135pp; English.  
 CC  
 CC The present invention relates to fusion proteins of Mycobacterium  
 CC tuberculosis antigens, and the polynucleotide sequences encoding them.  
 CC The sequences of the invention are useful in a method for preventing  
 CC tuberculosis by administering to a subject an amount of the fusion  
 CC protein or the polynucleotide that encodes the fusion protein. Also  
 CC disclosed is a pharmaceutical composition comprising the fusion protein  
 CC or the polynucleotide sequence encoding it. The fusion protein induces an  
 CC immune response to M. tuberculosis and can be used in the diagnosis,  
 CC prevention, and treatment of tuberculosis infection. The present sequence  
 CC represents a fusion protein of a M. tuberculosis antigen.  
 XX  
 SQ Sequence 368 AA;  
 Query Match 100.0%; Score 736; DB 7; Length 368;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-73;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MATTLPVQRHPRSLPPEFSELPAAPSPFAGLRPTFTRLMLRLEDEMKEGRYEVRALPGV 60  
 DB 9 MATTLPVQRHPRSLPPEFSELPAAPSPFAGLRPTFTRLMLRLEDEMKEGRYEVRALPGV 68  
 QY 61 DPDKVDIMVRDGLTIKARTEKQDPGRSEFAVGSFVRTVSLPVGADEDDIKATYDKG 120  
 DB 69 DPDKVDIMVRDGLTIKARTEKQDPGRSEFAVGSFVRTVSLPVGADEDDIKATYDKG 128  
 QY 121 ILTVSAVSEKGPTEKHIQIRSTN 144  
 DB 129 ILTVSAVSEKGPTEKHIQIRSTN 152  
 RESULT 10  
 AAY32065  
 ID AAY32065 standard; protein; 433 AA.  
 XX  
 AC AAY32065;  
 XX  
 DT 17-JAN-2000 (first entry)  
 XX  
 DE Mycobacterium tuberculosis antigen fusion protein Mtb46f.  
 XX  
 KM Tuberculosis; antigen; fusion protein; Mtb46f; ERD14; DPV; MTI; MSI;  
 KM diagnosis; therapy; vaccine; immunogen.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..8  
 FT Protein /note= "Met/His tag"  
 FT Protein /note= "ERD14"  
 FT Protein 155..236  
 FT Protein /note= "DPV"  
 FT Protein 239..332  
 FT Protein /note= "MTI"  
 FT Protein 335..433  
 FT Protein /note= "MSI"  
 PN MO9951748-A2.  
 PD 14-OCT-1999.  
 XX  
 PF 07-APR-1999; 99WO-US007717.  
 XX  
 PR 07-APR-1998; 98US-00056556.  
 PR 30-DEC-1998; 98US-00223040.  
 XX  
 PA (CORI-) CORIXA CORP.

XX  
 FI Skeiky YAM, Alderson M, Campos-Neco A;  
 FI WPI; 1999-601610/51.  
 DR N-PSDB; AAZ20200.  
 XX  
 PT New fusion proteins useful for diagnosis, prevention and treatment of  
 PT tuberculosis.  
 XX  
 PS Claim 1; Fig 7A-B; 83pp; English.  
 CC  
 CC This sequence represents a recombinant Mycobacterium tuberculosis tetra-  
 CC antigen fusion protein, termed Mtb46f, composed of the antigens ERD14,  
 CC DPV, MTI and MSI. The fusion protein is expressed in host cells using a  
 CC vector carrying a polynucleotide (see AA220200) comprising the coding  
 CC sequences for the 4 antigens. The invention provides fusion proteins (see  
 CC AA32059-71) containing at least 2 M. tuberculosis antigens. The new  
 CC fusion proteins and polynucleotides encoding them are useful as vaccines  
 CC for preventing tuberculosis (claimed), for diagnosis (via in vitro assays  
 CC or intradermal skin tests for detection of anti-M. tuberculosis  
 CC antibodies), monitoring of disease progression, and treatment of  
 CC tuberculosis. They are more effective immunogens than mixtures of the  
 CC individual protein components  
 XX  
 SQ Sequence 433 AA;  
 Query Match 100.0%; Score 736; DB 2; Length 433;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-73;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MATTLPVQRHPRSLPPEFSELPAAPSPFAGLRPTFTRLMLRLEDEMKEGRYEVRALPGV 60  
 DB 9 MATTLPVQRHPRSLPPEFSELPAAPSPFAGLRPTFTRLMLRLEDEMKEGRYEVRALPGV 68  
 QY 61 DPDKVDIMVRDGLTIKARTEKQDPGRSEFAVGSFVRTVSLPVGADEDDIKATYDKG 120  
 DB 69 DPDKVDIMVRDGLTIKARTEKQDPGRSEFAVGSFVRTVSLPVGADEDDIKATYDKG 128  
 QY 121 ILTVSAVSEKGPTEKHIQIRSTN 144  
 DB 129 ILTVSAVSEKGPTEKHIQIRSTN 152  
 RESULT 11  
 AAU74594  
 ID AAU74594 standard; protein; 433 AA.  
 XX  
 AC AAU74594;  
 XX  
 DT 29-AUG-2003 (revised)  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Antigenic fusion protein Erd14-DPV-MTI-MSI (Mtb46f).  
 XX  
 KM Fusion protein; tuberculosis; Mycobacterium tuberculosis;  
 KM tuberculostatic; immunogen; vaccine; Erd14-DPV-MTI-MSI; Mtb46f.  
 XX  
 OS Mycobacterium tuberculosis.  
 OS Chimeric.  
 FN US2002009459-A1.  
 XX  
 PD 24-JAN-2002.  
 XX  
 PF 07-APR-1999; 99US-00287849.  
 XX  
 PR 13-MAR-1997; 97US-00818112.  
 PR 01-OCT-1997; 97US-00942578.  
 PR 18-FEB-1998; 98US-00025197.  
 PR 07-APR-1998; 98US-00056556.  
 PR 30-DEC-1998; 98US-00223040.  
 XX  
 PA (REED/) REED S G.

PA (SKEI/) SKEIKY Y A.  
PA (DILL/) DILLON D C.  
PA (ALDE/) ALDERSON M.  
PA (CAMP/) CAMPOS-NETO A.  
PI Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;  
XX WPI; 2002-171134/22.  
DR N-PSDB; ABK14134.  
XX  
XX  
PT New fusion proteins of Mycobacterium tuberculosis antigens, useful for  
PT diagnosing, treating or preventing M. tuberculosis infection,  
PT particularly as vaccine for treating or preventing tuberculosis.  
XX  
XX  
PS Claim 1; Fig 7; 62pp; English.  
XX  
CC The invention relates to a purified polypeptide which induces an immune  
CC response of Mycobacterium tuberculosis. Polypeptides of the invention are  
CC useful for diagnosing, treating or preventing M. tuberculosis infection,  
CC particularly tuberculosis infection. In particular, the polypeptides are  
CC useful as a vaccine formulation with an adjuvant to afford long-term  
CC protection in animals against the development of tuberculosis. The  
CC protein coding sequence may be used to encode a protein product for use  
CC as an immunogen to induce and/or enhance an immune response to M.  
CC tuberculosis. This sequence represents an M. tuberculosis fusion protein  
CC of the invention. (Updated on 29-AUG-2003 to standardise OS field)  
XX  
SQ Sequence 433 AA;

Query Match 100.0%; Score 736; DB 5; Length 433;  
Best Local Similarity 100.0%; Pred. No. 3.9e-73;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATTLPVQRHRSLEPFSESLFAAFPSFAGLRPTFDTRLMLBDEMEKGRYEVAAELPGV 60  
DB 9 MATTLPVQRHRSLEPFSESLFAAFPSFAGLRPTFDTRLMLBDEMEKGRYEVAAELPGV 68  
QY 61 DPDXDVIWVRDQGLTIKARTEQKDPGRSEFAVGSFVRTVSLPVGADEDDIKATYDKG 120  
DB 69 DPDXDVIWVRDQGLTIKARTEQKDPGRSEFAVGSFVRTVSLPVGADEDDIKATYDKG 128  
QY 121 ILTVSAVSESGKPEKHIQIRSTN 144  
DB 129 ILTVSAVSESGKPEKHIQIRSTN 152

RESULT 12  
ADP69765  
ID ADP69765 standard; protein; 433 AA.

XX ADF69765;

XX 12-FEB-2004 (first entry)

XX Fusion protein #7 of M. tuberculosis antigen.

XX Fusion protein; Mycobacterium tuberculosis antigen;

XX tuberculosis infection; immune response; tuberculostatic; mutant; mutein.

XX Synthetic.

OS Mycobacterium tuberculosis.

XX US2003147911-A1.

XX 07-AUG-2003.

XX 05-FEB-2003; 2003US-00359460.

XX 13-MAR-1997; 97US-00818112.

XX 01-OCT-1997; 97US-00942578.

XX 18-FEB-1998; 98US-00025197.

XX 07-APR-1998; 98US-00056556.

XX 30-DEC-1998; 98US-00223040.

PR 07-APR-1999; 99US-00287849.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX  
PI Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;  
XX WPI; 2003-897524/82.  
DR N-PSDB; ADP69764.  
XX  
XX  
XX  
PT New fusion proteins of Mycobacterium tuberculosis for diagnosing,  
PT preventing or treating tuberculosis infection or in enhancing immune  
PT responses in M. tuberculosis.  
XX  
XX  
PS Claim 1; SEQ ID NO 14; 135pp; English.  
XX  
XX  
CC The present invention relates to fusion proteins of Mycobacterium  
CC tuberculosis antigens, and the polynucleotide sequences encoding them.  
CC The sequences of the invention are useful in a method for preventing  
CC tuberculosis by administering to a subject an amount of the fusion  
CC protein or the polynucleotide that encodes the fusion protein. Also  
CC disclosed is a pharmaceutical composition comprising the fusion protein  
CC or the polynucleotide sequence encoding it. The fusion protein induces an  
CC immune response to M. tuberculosis and can be used in the diagnosis,  
CC prevention, and treatment of tuberculosis infection. The present sequence  
CC represents a fusion protein of a M. tuberculosis antigen.  
XX  
SQ Sequence 433 AA;

Query Match 100.0%; Score 736; DB 7; Length 433;  
Best Local Similarity 100.0%; Pred. No. 3.9e-73;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATTLPVQRHRSLEPFSESLFAAFPSFAGLRPTFDTRLMLBDEMEKGRYEVAAELPGV 60  
DB 9 MATTLPVQRHRSLEPFSESLFAAFPSFAGLRPTFDTRLMLBDEMEKGRYEVAAELPGV 68  
QY 61 DPDXDVIWVRDQGLTIKARTEQKDPGRSEFAVGSFVRTVSLPVGADEDDIKATYDKG 120  
DB 69 DPDXDVIWVRDQGLTIKARTEQKDPGRSEFAVGSFVRTVSLPVGADEDDIKATYDKG 128  
QY 121 ILTVSAVSESGKPEKHIQIRSTN 144  
DB 129 ILTVSAVSESGKPEKHIQIRSTN 152

RESULT 13  
AAI32064  
ID AAI32064 standard; protein; 856 AA.

XX AAY32064;

XX 17-JAN-2000 (first entry)

XX Mycobacterium tuberculosis antigen fusion protein Mtb88f.

XX Tuberculosis; antigen; fusion protein; Mtb88f; ERD14; DPV; MTI; MSL;

XX mTCC2; diagnosis; therapy; vaccine; immunogen.

XX Mycobacterium tuberculosis.

OS Location/Qualifiers

XX Key 1..8

XX Peptide /note= "Met/His tag"

XX Protein 9..152

XX Protein /note= "ERD14"

XX Protein 153..236

XX Protein /note= "DPV"

XX Protein 239..332

XX Protein /note= "MTI"

XX Protein 335..431

XX Protein /note= "MSL"

XX Protein 434..856

XX Protein /note= "mTTC2"

```

XX XX MO9951748-A2.
XX PD 14-OCT-1999.
XX PF 07-APR-1999; 99WO-US007717.
XX PR 07-APR-1998; 98US-00056556.
XX PR 30-DEC-1998; 98US-00223040.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky YAM, Alderson M, Campos-Neto A;
XX DR WPI; 1999-601610/51.
XX DR N-PSDB; AA220199.
XX PT New fusion proteins useful for diagnosis, prevention and treatment of
XX PT tuberculosis.
XX PS Claim 1; Fig 6A-B; 83pp; English.
XX CC This sequence represents a recombinant Mycobacterium tuberculosis penta-
CC antigen fusion protein, termed Mtb88f, composed of the antigens Erd14,
CC DpV, MTI, MSU and mTTC2. The fusion protein is expressed in host cells
CC using a vector carrying a polynucleotide (see AA220199) comprising the
CC coding sequences for the 5 antigens. The invention provides fusion
CC proteins (see AA32059-71) containing at least 2 M. tuberculosis
CC antigens. The new fusion proteins and polynucleotides encoding them are
CC useful as vaccines for preventing tuberculosis (claimed), for diagnosis
CC (via in vitro assays or intradermal skin tests for detection of anti-M.
CC tuberculosis antibodies), monitoring of disease progression, and
CC treatment of tuberculosis. They are more effective immunogens than
CC mixtures of the individual protein components
XX SQ Sequence 856 AA;

Query Match 100.0%; Score 736; DB 2; Length 856;
Best Local Similarity 100.0%; Pred. No. 1e-72;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATTLPVQRHPRSLPPEFSELFAAPSPFAGLRPTPTDRLMLRLEDEMKEGRYEVRALPGV 60
DB 9 MATTLPVQRHPRSLPPEFSELFAAPSPFAGLRPTPTDRLMLRLEDEMKEGRYEVRALPGV 68
QY 61 DPDKVDIMVWDQGLTIKARTEQKDFDGRSEFAYGSFVRTVSLPVGADDDIKATYDKG 120
DB 69 DPDKVDIMVWDQGLTIKARTEQKDFDGRSEFAYGSFVRTVSLPVGADDDIKATYDKG 128
QY 121 ILTVSAVSESGKPTKEKHIIQIRSTN 144
DB 129 ILTVSAVSESGKPTKEKHIIQIRSTN 152

RESULT 14
AD69763
ID ADF69763 standard; protein; 856 AA.
XX
XX ADF69763;
AC
XX 12-FEB-2004 (first entry)
XX
XX Fusion protein #6 of M. tuberculosis antigen.
XX
XX Fusion protein; Mycobacterium tuberculosis antigen;
KM tuberculosis infection; immune response; tuberculostatic; mutant; munein.
XX
XX Synthetic.
OS Mycobacterium tuberculosis.
XX
XX US2003147911-A1.
XX
XX 07-AUG-2003.
PD

```

```

XX PF 05-FEB-2003; 2003US-00359460.
XX PR 13-MAR-1997; 97US-00818112.
XX PR 01-OCT-1997; 97US-00942578.
XX PR 18-FEB-1998; 98US-00025197.
XX PR 07-APR-1998; 98US-00056556.
XX PR 30-DEC-1998; 98US-00223040.
XX PR 07-APR-1999; 99US-00287849.
XX PA (CORI-) CORIXA CORP.
XX PI Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
XX DR WPI; 2003-897524/82.
XX DR N-PSDB; ADF69762.
XX PT New fusion proteins of Mycobacterium tuberculosis for diagnosing,
XX PT preventing or treating tuberculosis infection or in enhancing immune
XX PT responses in M. tuberculosis.
XX PS Claim 1; SEQ ID NO 12; 135pp; English.
XX CC The present invention relates to fusion proteins of Mycobacterium
XX CC tuberculosis antigens, and the polynucleotide sequences encoding them.
XX CC The sequences of the invention are useful in a method for preventing
XX CC tuberculosis by administering to a subject an amount of the fusion
XX CC protein or the polynucleotide that encodes the fusion protein. Also
XX CC disclosed is a pharmaceutical composition comprising the fusion protein
XX CC or the polynucleotide sequence encoding it. The fusion protein induces an
XX CC immune response to M. tuberculosis and can be used in the diagnosis,
XX CC prevention, and treatment of tuberculosis infection. The present sequence
XX CC represents a fusion protein of a M. tuberculosis antigen.
XX SQ Sequence 856 AA;

Query Match 100.0%; Score 736; DB 7; Length 856;
Best Local Similarity 100.0%; Pred. No. 1e-72;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATTLPVQRHPRSLPPEFSELFAAPSPFAGLRPTPTDRLMLRLEDEMKEGRYEVRALPGV 60
DB 9 MATTLPVQRHPRSLPPEFSELFAAPSPFAGLRPTPTDRLMLRLEDEMKEGRYEVRALPGV 68
QY 61 DPDKVDIMVWDQGLTIKARTEQKDFDGRSEFAYGSFVRTVSLPVGADDDIKATYDKG 120
DB 69 DPDKVDIMVWDQGLTIKARTEQKDFDGRSEFAYGSFVRTVSLPVGADDDIKATYDKG 128
QY 121 ILTVSAVSESGKPTKEKHIIQIRSTN 144
DB 129 ILTVSAVSESGKPTKEKHIIQIRSTN 152

RESULT 15
AAU74593
ID AAU74593 standard; protein; 859 AA.
XX
XX AAU74593;
AC
XX 29-AUG-2003 (revised)
XX DT 08-MAY-2002 (first entry)
XX
XX Antigenic fusion protein Erd14-DpV-MTI-MSU-MTCC2 (Mcb88f).
XX
XX Fusion protein; tuberculosis; Mycobacterium tuberculosis;
KM tuberculostatic; immunogen; vaccine; Mtb88f; Erd14-DpV-MTI-MSU-MTCC2.
XX
XX Mycobacterium tuberculosis.
OS Chimeric.
XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 857 /label= OTHER
PD

```

FT /note= "OTHER= Xaa, Xaa= In frame stop codon"

XX US2002009459-A1.

XX 24-JAN-2002.

XX 07-APR-1999; 99US-00287849.

XX 13-MAR-1997; 97US-00818112.

PR 01-OCT-1997; 97US-00942578.

PR 18-FEB-1998; 98US-00025197.

PR 07-APR-1998; 98US-00056556.

PR 30-DEC-1998; 98US-00223040.

XX (REED/) REED S G.

PA (SKEI/) SKEIKY Y A.

PA (DILL/) DILLON D C.

PA (ALDE/) ALDERSON M.

PA (CAMP/) CAMPOS-NETO A.

PI Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;

XX WPI; 2002-171134/22.

DR N-PSDB; ABK14133.

XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for

PT diagnosing, treating or preventing M. tuberculosis infection,

PT particularly as vaccine for treating or preventing tuberculosis.

XX Claim 1; Fig 6; 62pp; English.

XX The invention relates to a purified polypeptide which induces an immune

CC response of Mycobacterium tuberculosis. Polypeptides of the invention are

CC useful for diagnosing, treating or preventing M. tuberculosis infection,

CC particularly tuberculosis infection. In particular, the polypeptides are

CC useful as a vaccine formulation with an adjuvant to afford long-term

CC protection in animals against the development of tuberculosis. The

CC protein coding sequence may be used to encode a protein product for use

CC as an immunogen to induce and/or enhance an immune response to M.

CC tuberculosis. This sequence represents an M. tuberculosis fusion protein

CC of the invention. (updated on 29-AUG-2003 to standardise OS field)

XX Sequence 859 AA;

XX Query Match 100.0%; Score 736; DB 5; Length 859;

XX Best Local Similarity 100.0%; Pred. No. 1e-72; Mismatches 0; Gaps 0;

XX Matches 144; Conservative 0; Indels 0; Gaps 0;

QY 1 MATTLPVQRHPRSLPEFSESLFAAFSPFAGLPTFTRLMRLEDEMKEGRYEVRALPGV 60

Db 9 MATTLPVQRHPRSLPEFSESLFAAFSPFAGLPTFTRLMRLEDEMKEGRYEVRALPGV 68

QY 61 DPDKVDIMVDRDGLTIKARTKQKDPGRSFPAYGSFVRTVSLPVGADEDDIKATYDKG 120

Db 69 DPDKVDIMVDRDGLTIKARTKQKDPGRSFPAYGSFVRTVSLPVGADEDDIKATYDKG 128

QY 121 ILTVSVANSEKGPTEKHQIRSTN 144

Db 129 ILTVSVANSEKGPTEKHQIRSTN 152

Search completed: March 23, 2006, 05:10:46

Job time : 40.2848 secs



***This Page Blank (uspto)***

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 05:11:18 ; Search time 5.72908 Seconds  
(without alignments)  
2418.401 Million cell updates/sec

Title: US-10-617-038-24

Perfect score: 736

Sequence: 1 MATTLPVQRHPRSLPFSEFSE.....SVAVSEGGPKTEKHQIRSTN 144

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR\_80:\*

2: PIR1:\*

3: PIR2:\*

4: PIR3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	736	100.0	144	2	F70942 probable hsp90 prot
2	184.5	25.1	175	2	E55337 probable hsp90 hea
3	164.5	22.4	159	2	G70939 probable hsp prote
4	156	21.2	146	2	S74956 spore protein sp21
5	155.5	21.1	143	2	T40376 heat shock protein
6	151	20.5	168	2	P95941 probable small hea
7	150.5	20.4	188	2	A49942 heat shock protein
8	148	20.1	140	2	B69496 small heat shock p
9	140.5	19.1	151	2	T05739 probable heat shoc
10	139	18.9	160	2	F82582 low molecularweig
11	137.5	18.7	142	2	T46568 small heat shock p
12	136.5	18.5	152	2	S72546 heat shock protein
13	136.5	18.5	157	2	S04939 heat shock 22K pro
14	136	18.5	159	2	S72544 heat shock protein
15	133	18.1	151	2	S21600 heat shock protein
16	133	18.1	182	2	A75436 heat shock protein
17	132	17.9	147	2	D72385 heat shock protein
18	132	17.9	214	2	S45465 heat shock protein
19	132	17.9	229	2	S58210 heat shock protein
20	131.5	17.9	150	2	S72545 heat shock protein
21	131.5	17.9	149	2	T05740 heat shock protein
22	131	17.8	160	2	T04171 heat shock protein
23	130.5	17.7	159	1	CYP279 heat shock protein
24	130	17.7	151	1	HHMT17 heat shock protein
25	130	17.7	210	2	T09611 heat shock protein
26	129.5	17.6	150	2	A18142 small heat shock p
27	128	17.4	152	2	S23212 heat shock protein
28	128	17.4	153	2	S23529 heat shock protein
29	128	17.4	156	2	S71566 heat shock protein

30	127	17.3	232	2	C96566 hypothetical prote
31	126.5	17.2	154	2	S24396 heat shock protein
32	126.5	17.2	161	2	S00646 heat shock protein
33	126	17.1	157	2	T14381 heat-shock protein
34	126	17.1	160	2	TC4377 low-molecularweig
35	125.5	17.1	154	2	T44801 heat shock protein
36	125.5	17.1	155	2	T06449 probable heat shoc
37	125.5	17.1	161	2	S33566 heat shock protein
38	123.5	16.8	154	2	JS0710 heat shock protein
39	123.5	16.8	159	2	T04173 heat shock protein
40	123	16.7	157	2	S06074 heat shock protein
41	122.5	16.6	156	2	UQ0351 heat shock protein
42	122.5	16.6	161	2	UQ0352 heat shock protein
43	122.5	16.6	163	2	S71567 small heat-shock p
44	122	16.6	143	2	A57270 heat shock protein
45	121.5	16.5	150	2	S20874 heat shock protein

ALIGNMENTS

RESULT 1

F70942 probable hsp90 protein - Mycobacterium tuberculosis (strain H37Rv)  
N:Alternate names: 14K antigen; 19K major membrane protein  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence revision 17-Jul-1998 #text\_change 09-Jul-2004  
C:Accession: F70942; A42651; A43823

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holtz, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A: Authors: Sgares, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A: Reference number: A70500; MUID:9825987; PMID:9634230

A: Accession: F70942  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA  
A: Residues: 1-144 <COL>

A: Cross-references: UNIPROT:P30223; UNIPARC:UPI000004A4BF; GB:AL021899; GB:AL123456; NID  
A: Experimental source: strain H37Rv  
A: Verdon, A.; Hartkeel, R.A.; Schultema, A.; Kolk, A.H.; Young, D.B.; Lathigra, R.  
J. Bacteriol. 174, 1352-1359, 1992

A: Title: The 14,000-molecular-weight antigen of Mycobacterium tuberculosis is related to  
A: Reference number: A42651; MUID:92138631; PMID:1370952  
A: Accession: A42651  
A: Status: preliminary

A: Molecule type: DNA  
A: Residues: 1-144 <VER>

A: Cross-references: UNIPARC:UPI000004A4BF; GB:S79751; NID:9244561; PIDN:AA21317.1; PID:  
A: Lee, B.Y.; Hefta, S.A.; Brennan, P.J.  
Infect. Immun. 60, 2066-2074, 1992

A: Title: Characterization of the major membrane protein of virulent Mycobacterium tubercu  
A: Reference number: A43823; MUID:9225631; PMID:1563797  
A: Accession: A43823  
A: Molecule type: protein

A: Residues: 2-144 <LEB>  
A: Cross-references: UNIPARC:UPI0000124E06  
A: Experimental source: strain Erdman

A: Gene: hsp90  
A: Family: alpha-crystallin-related small heat shock protein  
C: Keywords: membrane protein; surface antigen

Query Match 100.0%; Score 736; DB 2; Length 144;  
Best Local Similarity 100.0%; Pred. No. 4.8e-60;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MATTLPVQRHPRSLPFSEFSEFAGLRPTFTRLRLDEMEGRYVAAELPGV	60
Db	1	MATTLPVQRHPRSLPFSEFSEFAGLRPTFTRLRLDEMEGRYVAAELPGV	60



```

Beat Local Similarity    33.6%; Pred. No.1.2e-06; Mismatches 47; Indels 16; Gaps 6; Matches 45; Conservative 26;
QY      5 LPVGRHPSPFPEFSSELF-----AAPPSPAGLRPTFDTRLMLRLEDEMKRGREYVRALP 58
DB      29 LATHRMARLTFDPAFRSFETRLPLSGFSFGAGMPS-----VEVSDDKE--IKVTALRP 81
OY     59 GVDPDKVDIMWRDGLTIKAERTEOKDFGSEFYATSVLPVGADDEDIATK 116
DB      82 GID-EKVVELSLSDGVLSIRGEKRATEDENOFSERYGRFERRIALLGYVEDSKVAWT 140
QY      117 YDKGITLVSAVNSE 130
DB      141 FENGVLTVTLPKTE 154

RESULT 7
A49942
heat shock protein SPZ1 - Stigmatella aurantiaca
N:Alternates names: spore protein
C:Species: Stigmatella aurantiaca
C>Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 09-Jul-2004
C:Accession: A49942; S27669
R:Heidelbach, M.; Skladny, H.; Schaller, H.U.
J:Bacteriol. 175, 7479-7482, 1993
A>Title: Heat shock and development induce synthesis of a low-molecular-weight stress-resistance proteins from Streptomyces griseus
A:Reference number: A49942; PMID:822695
A:Accession: A49942
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-188 <HE>
A:Cross-references: UNIPROT:Q06823; UNIPARC:UPI0000135D57; GB:M94510; NID:g152657; PIDN::C:Genetics:
A:Gene: hspa
C:Superfamily: alpha-crystallin-related small heat shock protein

Query Match          20.4%; Score 150.5; DB 2; Length 188;
Best Local Similarity 27.9%; Pred. No.1.5e-06;
Matches 43; Conservative 29; Mismatches 45; Indels 37; Gaps 7;

QY      15 PFESEL-----FAFPSPAFLR-----PFEDTRLMLEDMEKGREYVRALPGVD 61
DB      23 FOQMQLMNWDPELLNNHFWFNRRGPRAVFAPAEVR-----ETKEA-YIFKDLPGVD 75
OY     62 PDKDVIDMRDGOLITIKAERTEQKDFGSRSEPAY---GSFVRTVSLPVGADEDDIKATY 117
DB      76 EKDIIEVLTIGRVAVSNGRKEREKERESRFVYKRTFGSFRAFLLPEGVGDGNVRADL 134
QY      118 DKGITLVSV-----AVSECKPTEKHIIQ 139
DB      135 KNGVLTVTLTPKKREVOPKRIQVASSTGEQKEHIK 168

RESULT 8
B69496
small heat shock protein (hsp20-2) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C:Accession: B69496
R:Klikentz, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
A.: Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes, S.I.; Smith, H.O.; Woese, C.R.; Venter, J.C.
A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Pyrococcus furiosus
A:Reference number: A69250; PMID:98049343; PMID:9389475
A:Accession: B69496
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-140 <KL>-
A:Cross-references: UNIPROT:Q28308; UNIPARC:UPI0000056B10; GB:AEO00967; GB:AEO00782; NID
C:Superfamily: alpha-crystallin-related small heat shock protein
```

```

Query Match      20.1%; Score 148; DB 2; Length 140;
Best Local Similarity 32.8%; Pred. No. 1.8e-06;
Matches 45; Conservative 20; Mismatches 58; Indels 14; Gaps 6;

OY  RHPRLSPFESGLTFAFPSPFAGLRPTFDTRLMLRLDEMKEG-RYEVRAELPGVDPDKVD 67
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB  RMRGRFRRLLEEFGRFPEVKRFTVMPY-----DVDEGEQIRVAVDLDFGSKK-DLE 64
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

OY  IMVRDQGLTIKAERTQ-KDFDG--RSEFAYGSFVRTVSLPGVADDEDIKATYDKGILT 123
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB  IYFEDGDLVIAKAEKKEEFEEKGEYLRREBRMGKYRRIALPAGHDIDAVAKYNNGVL- 123
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

OY  124 VSAVSEKPTREKHQI 140
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB  124 -EITPKLKDKRKAVQI 139
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 9
T05739
probable heat shock protein 17 - barley
C|Species: Hordeum vulgare (barley)
C|Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 09-Jul-2004
C|Accession: T05739
R|Marzocchi, N.; Pavoni, A.; di Cola, G.; Hartings, H.; Raho, G.; Conte, M.R.; Petrotta,
Genome 36, 1111-1118, 1993
A|Title: Identification, characterization and analysis of cDNA and genomic sequences enc
A|Reference number: Z15446; MUID:94156165; PMID:8112573
A|Accession: T05739
A|Status: preliminary; translated from GB/EMBL/DBJ
A|Molecule type: DNA
A|Residues: 1-151 <MAR>
A|Cross-references: UNIPROT:Q40056; UNIPARC:UP100000A2508; EMBL:X64560; NID:G509175; PIR
A|Experimental source: cv. Onice
C|Genetics:
A|Gene: hsp17
C|Superfamily: alpha-crystallin-related small heat shock protein
C|Keywords: heat shock; stress-induced protein

Query Match      19.1%; Score 140.5; DB 2; Length 151;
Best Local Similarity 32.7%; Pred. No. 9.6e-06;
Matches 50; Conservative 25; Mismatches 57; Indels 21; Gaps 8;

OY  5 LPVGRHPRSLPPSPSELPF-----AFSPAGLRPTFDTRLMLRLDEMKEGRYEV--- 53
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB  1 MSIVRSNVLPD-PADLMADPLDTFRSLTFPAISGNSETAVERRRM--DWKGRLEAHVF 57
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

OY  54 RAELPGVDPRKDVIDMVRDQ-LTIKAERTQKDFG---RSEFAYGSFVRTVSLPGVA 108
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB  58 KADLPGVKKE-EVKKEVEDGNVLIVSGERTKEKDKNDKMHVERBSGKFPVRPLRPEDG 116
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

OY  109 DEDIKATYDKGILTVSAVSE-GKPTREKHQI 140
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB  117 KYDEVKAGLENGVLTVTVPKAEVKKPEVKALEI 149
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 10
F82582
low molecular weight heat shock protein Xr2234 [imported] - Xylella faacidiosa (strain 5
C|Species: Xylella faacidiosa
C|Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C|Accession: F82582
R|Anonymous, The Xylella faacidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A|Title: The genome sequence of the plant pathogen Xylella faacidiosa.
A|Reference number: A82515; MUID:20365117; PMID:10910347
A|Note: for a complete list of authors see reference number A59328 below
A|Accession: F82582
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-160 <STM>
A|Cross-references: UNIPROT:Q9PB80; UNIPARC:UP100000C239B; GB:AE004036; GB:AE003849; NID
A|Experimental source: strain 9A5C

```

R. Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A. Britones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter, H. as-Neco, E.; Docena, C.; El-Dorfy, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fromme, J.D.; Jungueira, M.L.; Kemper, E.L.; Kitaajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, A.; Chado, M.A.; Madeira, M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E. A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasakki, M.A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, M.; Teshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z...

A:Reference number: A59328  
C:Contents: annotation  
A:Gene: XF2234  
C:Superfamily: alpha-crystallin-related small heat shock protein

Query Match 18.9%; Score 139; DB 2; Length 160;  
Best Local Similarity 32.7%; Pred. No. 1,4e-05;  
Matches 34; Conservative 24; Mismatches 40; Indels 6; Gaps 3;

Qy 41 RIIDEMKEGRYEVRAELPGVDPDKVDYDINWRDQLTKAER---TEQKDFDGRSEFAFG 96  
Db 46 RVDIKENPQFVLYADLPGLDPA-ADIEVQMDKILSLKGRKTBSSQTHFSRIERRYG 104

Qy 97 SEFRTVSLPVGADDEDIDIKATYDKGILTVSAVASEGKPTKHIQI 140  
Db 105 SFHRRPALPDSADADGITASGSHGVLSIFIP-KRAATTPTRIQV 147

RESULT 11  
T46658  
small heat shock protein [validated] - Thermotoga maritima (DSM 3109)  
C:Species: Thermotoga maritima  
A:Variety: DSM 3109 #sequence\_revision 18-Feb-2000 #text\_change 09-Jul-2004  
C:Date: 18-Feb-2000  
C:Accession: T46658  
R.Michellini, E.T.; Flynn, G.C.  
submitted to the EMBL Data Library, November 1998  
A:Description: The unique chaperone operon of Thermotoga maritima: cloning and initial ch  
A:Reference number: Z23119  
A:Accession: T46658  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-142 <MIC-  
A:Cross-references: UNIPROT:Q9ZFP1; UNIPARC:UP1000000BDB94; EMBL:AF106330; PDB:AACT9726.1  
A:Experimental source: DSM 3109  
C:Genetics:  
A:Gene: shsp  
C:Function:  
A:Description: molecular chaperone [validated, MUID:99328968]  
C:Superfamily: alpha-crystallin-related small heat shock protein

Query Match 18.7%; Score 137.5; DB 2; Length 142;  
Best Local Similarity 29.7%; Pred. No. 1.7e-05;  
Matches 41; Conservative 25; Mismatches 53; Indels 19; Gaps 5;

Qy 7 VQHRHRSLPFSESELPAAPFAGLAPFTDRLMRLEDEMKEGRYEVRAELPGVDPDKDV 66  
Db 16 LQREIDRLPDDFFR-----TEVRPAPMDVDFETDDEV-----VLEVIPLGID-RKDV 61

Qy 67 DIMWRDQGLTIKAERTQKDFDGRS---EFAVGSFVRTSLPVGADDEDIDIKATYDKGIL 122  
Db 62 QITVEENLIKIGSEKLEBQKKNYYVERSGKFERAIRLPDYDVVEKIKAEYKNGVL 121

Qy 123 TVSAVASEGKPTKHIQI 140  
Db 122 TIRVPKKEER-KKQVIEV 136

RESULT 12  
S72546

heat shock protein 17.0 - pearl millet

C:Species: Pennisetum americanum (pearl millet)

C>Date: 14-Apr-1998 #sequence\_revision 15-May-1998 #text\_change 09-Jul-2004

C:Accession: S72546

R:Cavan, G.P.; Skoet, K.; Stevens, M.J.; Howarth, C.J.

A:Title: Sequence announcement.

A:Reference number: S72545

A:Accession: S72546

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 1-152 <CAV>

A:Cross-references: UNIPROT:Q40866; UNIPARC:UPI00000DD14; EMBL:X94191; NID:g1122314; PI

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1995

A>Note: the source is designated as Pennisetum glaucum

C:Superfamily: alpha-crystallin-related small heat shock protein

Query Match 18.5%; Score 136.5; DB 2; Length 152;

Best Local Similarity 32.4%; Pred. No. 2.2e-05;

Matches 44; Conservative 24; Mismatches 57; Indels 11; Gaps 6;

14 LFPFSESLFAFPSPAGLRPTFDRLM---RLDEMEKGRYEVRAELPGVDPKVDIMV 70

17 LWDPPDSMFRSIVSAG-SPSDTAAPAAARIDMKETPEAHVFKAADLPGVKKE-EVKVEV 74

71 RDGQ-LTIKARTEQKDFG---RSEFAYGSFVTVSLPVGADDDIKATYDKGILTVS 125

75 EDGNVLVLSGGRSKKEDKNDKMKHVERSSGGPFMRFLPNKAKVDQYKAGLNGVLTIVT 134

126 V-AVSEGGKPTKHIQI 140

135 VPKAEKKKEVKAIEI 150

Db

RESULT 13

heat shock 22K protein - Chlamydomonas reinhardtii

C:Species: Chlamydomonas reinhardtii

C>Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 09-Jul-2004

C:Accession: S04939

R:Griffith, B.; Ish-Shalom, D.; Even, D.; Glaczinski, H.; Otterbach, P.; Ohad, I.; Kloppe

Eur. J. Biochem. 182, 539-546, 1989

A:Title: The nuclear-coded chloroplast 22-kDa heat-shock protein of Chlamydomonas. Evid

A:Reference number: S04939; MUID:89325298; PMID:2473899

A:Accession: S04939

A:Molecule type: mRNA

A:Residues: 1-157 <GRI>

A:Cross-references: UNIPROT:P12811; UNIPARC:UPI000012C15; EMBL:X15053; NID:g18151; PIDN

C:Superfamily: alpha-crystallin-related small heat shock protein

C:Keywords: chloroplast

Query Match 18.5%; Score 136.5; DB 2; Length 157;

Best Local Similarity 38.8%; Pred. No. 2.3e-05;

Matches 38; Conservative 14; Mismatches 37; Indels 9; Gaps 4;

51 YEVAELRGVDPDQVDIMVDRGQLTIKARL---EQKDFG---RSEFAYGSFVTVSL 104

58 FELHDAAGWGD-DVKVLEGLVMTVGERKLSHTTBAGGKTKWRSRTAVSFRAFL 116

105 PVGADDDIKATYDKGILTVSAVSE--GKPTKHIQI 140

117 PENANPDGITAMDKGVLVTVPRKRPPAKPEPKIAIV 154

Db

RESULT 14

heat shock protein 17.9 - pearl millet

C:Species: Pennisetum americanum (pearl millet)

C>Date: 01-May-1998 #sequence\_revision 01-May-1998 #text\_change 09-Jul-2004

C:Accession: S72544

R:Cavan, G.P.; Skoet, K.; Stevens, M.J.; Howarth, C.J.

A:Title: Sequence announcement.

A:Reference number: S72544

A:Accession: S72544

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 1-152 <CAV>

A:Cross-references: UNIPROT:Q40867; UNIPARC:UPI00000AAAF7; EMBL:X94193; NID:g1122316; PI

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1995

A>Note: the source is designated as Pennisetum glaucum

C:Superfamily: alpha-crystallin-related small heat shock protein

A:Reference number: S72544

A:Accession: S72544

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 1-159 <CAV>

A:Cross-references: UNIPROT:Q40867; UNIPARC:UPI00000AAAF7; EMBL:X94193; NID:g1122316; PI

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1995

A>Note: the source is designated as Pennisetum glaucum

C:Genetics: hsp17.9

C:Superfamily: alpha-crystallin-related small heat shock protein

C:Keywords: heat shock; stress-induced protein

Query Match 18.5%; Score 136; DB 2; Length 159;

Best Local Similarity 32.8%; Pred. No. 2.6e-05;

Matches 44; Conservative 21; Mismatches 53; Indels 16; Gaps 5;

13 SLFPFSESLFAFPSPAGLRPTFDRLMRLDEMEKGRYEVRAELPGVDPKVDIMVDR 72

34 SLFPSPRTSSETAFAFG-----ARIDMKETPEAHVFKAADVPALKKE-EVKVEVD 83

73 GQ-LTIKARL---EQKDFGSEFAYGSFVTVSLPVGADDDIKATYDKGILTVSVA 127

84 GNVLIQISGERKKEEKTDTWHRVERSSGKPFRRFLPENAKTQIRASMEGVLTVTV 143

128 VSE-GKPTKHIQI 140

144 KEVKKPEVKSIOI 157

Db

RESULT 15

heat shock protein 16.9B - wheat

C:Species: Triticum aestivum (common wheat)

C>Date: 08-Jun-1994 #sequence\_revision 12-May-1995 #text\_change 09-Jul-2004

C:Accession: S21600

R:Weng, J.; Wang, Z.F.; Nguyen, H.T.

submitted to the EMBL Data Library, February 1992

A:Description: Cloning and characterization of cytoplasmic LHM HSP genes from wheat.

A:Reference number: S21600

A:Accession: S21600

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-151 <MEN>

A:Cross-references: UNIPROT:Q41560; UNIPARC:UPI00009F825; EMBL:X64618; NID:g21804; PIDN

C:Superfamily: alpha-crystallin-related small heat shock protein

C:Keywords: alpha-crystallin-related small heat shock protein

Query Match 18.1%; Score 133; DB 2; Length 151;

Best Local Similarity 31.5%; Pred. No. 4.6e-05;

Matches 45; Conservative 29; Mismatches 53; Indels 16; Gaps 7;

13 SLFPFSESLFAFPSPAGLRPTF-----DTRLM---RLDEMEKGRYEVRAELPGVD 63

8 NVFPDFPLWMDPDTFRSIVPAISGGSETAFAANARMDKETPEAHVFKAADLPGVKKE 67

64 KDQVIMRDPGQ-LTIKARTEQKDFG---RSEFAYGSFVTVSLPVGADDDIKATYD 118

68 -EVKVEVDGNVLVSGERTKEKDKNDKMKHVERSSGKPFRRFLLEDKVEEVKAGLE 126

119 KGILTVSAVSE-GKPTKHIQI 140

127 NGVLTIVPRKAEVKKPEVKAIOI 149

Db

Search completed: March 23, 2006, 05:26:30

Job time : 5.72908 secs

This Page Blank (uspto)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using SW model

Run on: March 23, 2006, 04:56:43 ; Search time 36.8238 Seconds  
(without alignments)  
2758.529 Million cell updates/sec

Title: US-10-617-038-24

Perfect score: 736  
Sequence: 1 MATTLPVGRHRSLLFPEFSE.....SVAVSEGGTEKHIQIRSTN 144

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_05.80.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	731	99.3	143	1 14KD_MYCBO	POA5b8 mycobacteri
2	731	99.3	143	1 14KD_MYCTU	POA5b7 mycobacteri
3	442	60.1	86	2 08KY85_MYCTU	08KY85 mycobacteri
4	442	60.1	86	2 08KY81_MYCS2	08KY81 mycobacteri
5	442	60.1	86	2 08KY73_MYCTT	08KY73 mycobacteri
6	442	60.1	86	2 08KY82_MYCGO	08KY82 mycobacteri
7	442	60.1	86	2 08KY70_MYCO	08KY70 mycobacteri
8	442	60.1	86	2 08KY76_MYCGN	08KY76 mycobacteri
9	442	60.1	86	2 08KY72_MYCC	08KY72 mycobacteri
10	441	59.9	86	2 08KY68_MYCH	08KY68 mycobacteri
11	441	59.9	86	2 08KY75_MYCPA	08KY75 mycobacteri
12	434	59.0	86	2 08KY69_MYCFO	08KY69 mycobacteri
13	363	49.3	86	2 08KY67_MYCXE	08KY67 mycobacteri
14	343	46.6	86	2 08KY84_MYCNO	08KY84 mycobacteri
15	343	46.6	86	2 08KY71_MYCGS	08KY71 mycobacteri
16	340	46.2	86	2 08KY80_MYCNR	08KY80 mycobacteri
17	340	46.2	86	2 08KY79_MYCUL	08KY79 mycobacteri
18	336	45.7	86	2 08KY78_MYCUL	08KY78 mycobacteri
19	333	45.2	86	2 08KY77_MYCUL	08KY77 mycobacteri
20	235.5	32.0	139	2 04NAQ3_SMICC	04NAQ3 mycobacte
21	232	31.5	157	2 04NM99_SMICC	04NM99 mycobacte
22	188	25.5	148	2 08KY74_MYCHA	08KY74 mycobacteri
23	187	25.4	166	2 08NS55_RHOA	08NS55 rhodopside
24	184.5	25.1	175	2 092287_RHIME	092287 rhizobium m
25	184	25.0	48	2 08KHQ1_MYCUL	08KHQ1 mycobacteri
26	184	25.0	48	2 08KY83_MYCUL	08KY83 mycobacteri
27	173	23.5	173	2 089LA9_BRASA	089LA9 bradyrhizob
28	171.5	23.3	147	2 074FR8_GEOSL	074FR8 geobacter 8
29	170	23.1	146	2 073TL7_MYCPA	073TL7 mycobacteri
30	166.5	22.6	231	2 088HW9_PSEPK	088HW9 pseudomonas
31	165.5	22.5	154	2 04UKR8_RICFE	04UKR8 rickettsia

32	165	22.4	198	2	041YM3_AZOVI	041YM3 azotobacter
33	164.5	22.4	197	2	06MJP8_EDEBA	06MJP8 bdellovibri
34	164.5	22.4	159	2	07U2G3_MYCBO	07U2G3 mycobacteri
35	164.5	22.4	159	2	053673_MYCTU	053673 mycobacteri
36	161.5	21.9	189	2	05K552_9BACT	05K552 uncultured
37	161	21.9	142	2	05NE64_FRATT	05NE64 francisella
38	160.5	21.8	166	2	04TIT0_GSPRN	04TIT0 erythrobact
39	158.5	21.5	158	2	04HAV5_9DBIO	04HAV5 deinococcus
40	158.5	21.5	188	2	052TH0_LEGPM	052TH0 legionella
41	158	21.5	145	2	09XDG8_MYCSM	09XDG8 mycobacteri
42	156	21.2	146	2	P72977_SYNXX	P72977 synecocyst
43	155.5	21.1	140	2	06ACP4_LEIXX	06ACP4 leifsonia x
44	155.5	21.1	143	1	HSP16_SCHPO	014368 schistosach
45	155.5	21.1	151	2	07UF81_RHOBA	07UF81 rhodopirell

## ALIGNMENTS

```

RESULT 1
ID 14KD_MYCBO STANDARD; PRT; 143 AA.
AC POA5B8; P30223;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE 14 kDa antigen (16 kDa antigen) (HSP 16.3).
GN Name=hspX; OrderedLocustNames=Mb2057c;
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacterium;
OC Mycobacterium tuberculosis complex.
OX NCBI_TaxId=1765;
RN [1]
RP NUCLEOTIDE SEQUENCE (LARGE SCALE GENOMIC DNA).
RC STRAIN=AF2122/97; PubMed=12788972; DOI=10.1073/pnas.1130426100;
RX MEDLINS=22709107;
RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthey S., Grondin S., Lacroix C., Monempe C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewison R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -1- SUBCELLULAR LOCATION: Probably the external side of the cell wall
CC -1- (By similarity).
CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)
CC family.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; BX248341; CAD96910.1; -; genomic_DNA.
CC InterPro; IPR002068; HSP20.
CC Pfam; PF00011; HSP20; 1.
CC PROSITE; PS01031; HSP20; 1.
CC Antigen; Complete proteome.
CC INTACT; 0
CC SEQUENCE 143 AA; 16096 MW; 751AD94203226C9 CRC64;
Query March 99.3%; Score 731; DB 1; Length 143;
Best Local Similarity 100.0%; Pred. No. 2e-58;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 ATTLPVGRHRSLLFPEFSELPFAFPSPFAGLPFFDTRLMKLEDEMKEGRYEVAELPGVD 61
DB 1 ATTLPVGRHRSLLFPEFSELPFAFPSPFAGLPFFDTRLMKLEDEMKEGRYEVAELPGVD 60
QY 62 PDKVDIMVDDGQITTAERTEQDPGRSEFAYGSVVRVYSLPVGDEDDIDATYKGI 121
DB 61 PDKVDIMVDDGQITTAERTEQDPGRSEFAYGSVVRVYSLPVGDEDDIDATYDKGI 120

```



Qy	122	LTVSAVSEGKPTKHIQIRSTN	144
Db	121	LTVSAVSEGKPTKHIQIRSTN	143

## RESULT 2

14KD_MYCTU	STANDARD:	PRT:	143 AA.
ID_14KD_MYCTU			
AC	POA5E7; P30223;		
DT	01-JUL-1993 (Rel. 25, Created)		
DT	01-JUL-1993 (Rel. 26, Last sequence update)		
DT	10-MAY-2005 (Rel. 47, Last annotation update)		
DE	14 kDa antigen (16 kDa antigen) (HSP 15.3.)		
GN	Name=hspX; Order=rellocusName=rv2031c, MT2090; ORFNames=MTV018.18c		
OS	Mycobacterium tuberculosis.		
OS	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium;		
OC	Mycobacterium tuberculosis complex.		
OX	NCBI_TaxId=1773;		

CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use as long as its content is in no way modified and this statement is not  
CC removed.

DR EMBL, S79751; AAB21317.1; -, Genomic DNA.  
DR EMBL, M76712; AAA25342.1; -, Genomic DNA.  
DR EMBL, BX842578; CAA17245.1; -, Genomic DNA.  
DR EMBL, AE007059; AAK46369.1; -, Genomic DNA.  
DR PIR, F70942; F70942.  
DR TIGR, MT2090; -.  
DR Tuberculae; Rv2031c; -.  
DR InterPro; IPR002068; Hsp20.  
DR Pfam; PF00011; HSP20\_1.  
DR PROSITE; PS01031; HSP20\_1.  
KW Antigen; Complete proteome; Direct protein sequencing  
FT INIT MET 0 0  
SQ SEQUENCE 143 AA; 16096 MW; 751AD94203226CE9 CRC64

RESULT 3	
Q8KY85_MYCTU	
ID Q8KY85_MYCTU PRELIMINARY;	PRT; 86 AA

DR	EMBL: AF53444; AA696206.1;	Genomic_DNA.
DR	InterPro: IPR003068; Hsp20.	
DR	Pfam: PF00011; HSP20.1.	
DR	PROSITE: PS01031; HSP20; 1.	
FT	NON_TER	1
FT	NON_TER	86
QO	SEQUENCE	86 AA; 9814 MW; 1D6366CC0BDE5845B CRC64;

Qy 88 DGRSEFAYGSFVRTVSLPVGADDDI 113  
61 DGRSEFAYGSFVRTVSLPVGADDDI 86

## RESULT 4

ID Q8KY81\_MYCSZ PRELIMINARY; PRT; 86 AA.  
AC Q8KY81;  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Alpha-crystallin protein (Fragment).  
GN Name=acr;  
OS Mycobacterium szulgai.  
OC Bacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1787;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=NTC 10831;  
RA Shampura I.C., Bastian I., Fonteyne P.-A., Portaele F.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20) family.  
DR EMBL; AF253448; AAM69210.1; -; Genomic\_DNA.  
DR InterPro; IPR002068; Hsp20.  
DR Pfam; PF00011; HSP20; 1.  
DR PROSITE; PS01031; HSP20; 1.  
FT NON\_TER 1  
SQ SEQUENCE 86 AA; 9814 MW; 1D6966C08D5845B CRC64;

Query Match 60.1%; Score 442; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.9e-32;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 28 FAGLRPTFDTRLMLRLEDEMKRGYEVRAELPGVDPDKVDIMVRDGLTTIAERTQKDF 87  
Db 1 FAGLRPTFDTRLMLRLEDEMKRGYEVRAELPGVDPDKVDIMVRDGLTTIAERTQKDF 60

Qy 88 DGRSEFAYGSFVRTVSLPVGADDDI 113  
61 DGRSEFAYGSFVRTVSLPVGADDDI 86

## RESULT 5

ID Q8KY73\_MYCIT PRELIMINARY; PRT; 86 AA.  
AC Q8KY73;  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Alpha-crystallin protein (Fragment).  
GN Name=acr;  
OS Mycobacterium intracellulare.  
OC Bacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacteriaceae; Mycobacterium;  
OC Mycobacterium avium complex (MAC).  
OX NCBI\_TaxID=1767;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=M4199;  
RA Shampura I.C., Bastian I., Fonteyne P.-A., Portaele F.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20) family.  
DR EMBL; AF253459; AAM69221.1; -; Genomic\_DNA.  
DR InterPro; IPR002068; Hsp20.  
DR Pfam; PF00011; HSP20; 1.  
DR PROSITE; PS01031; HSP20; 1.  
FT NON\_TER 1

FT NON\_TER 86  
SQ SEQUENCE 86 AA; 9814 MW; 1D6966C08D5845B CRC64;  
Query Match 60.1%; Score 442; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.9e-32;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 28 FAGLRPTFDTRLMLRLEDEMKRGYEVRAELPGVDPDKVDIMVRDGLTTIAERTQKDF 87  
Db 1 FAGLRPTFDTRLMLRLEDEMKRGYEVRAELPGVDPDKVDIMVRDGLTTIAERTQKDF 60

Qy 88 DGRSEFAYGSFVRTVSLPVGADDDI 113  
61 DGRSEFAYGSFVRTVSLPVGADDDI 86

## RESULT 6

ID Q8KY82\_MYCGO PRELIMINARY; PRT; 86 AA.  
AC Q8KY82;  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Alpha-crystallin protein (Fragment).  
GN Name=acr;  
OS Mycobacterium gordonae.  
OC Bacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1778;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=ATCC14470;  
RA Shampura I.C., Bastian I., Fonteyne P.-A., Portaele F.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20) family.  
DR EMBL; AF253447; AAM69209.1; -; Genomic\_DNA.  
DR InterPro; IPR002068; Hsp20.  
DR Pfam; PF00011; HSP20; 1.  
DR PROSITE; PS01031; HSP20; 1.  
FT NON\_TER 1  
SQ SEQUENCE 86 AA; 9814 MW; 1D6966C08D5845B CRC64;

Query Match 60.1%; Score 442; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.9e-32;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 28 FAGLRPTFDTRLMLRLEDEMKRGYEVRAELPGVDPDKVDIMVRDGLTTIAERTQKDF 87  
Db 1 FAGLRPTFDTRLMLRLEDEMKRGYEVRAELPGVDPDKVDIMVRDGLTTIAERTQKDF 60

Qy 88 DGRSEFAYGSFVRTVSLPVGADDDI 113  
61 DGRSEFAYGSFVRTVSLPVGADDDI 86

## RESULT 7

ID Q8KY70\_MYCO PRELIMINARY; PRT; 86 AA.  
AC Q8KY70;  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Alpha-crystallin protein (Fragment).  
GN Name=acr;  
OS Mycobacterium lentiflavum.  
OC Bacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacteriaceae; Mycobacterium;  
OX NCBI\_TaxID=141349;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=M96-1701;

Qy 88 DGRSEFAYGSFVRTVSLPVGADDDI 113  
61 DGRSEFAYGSFVRTVSLPVGADDDI 86

RA Shampula I.C., Baetian I., Fonteyne P.-A., Portaeis F.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)  
family.  
DR EMBL: AF253462; AAM69224.1; -; Genomic\_DNA.  
DR InterPro: IPR002068; Hsp20.  
DR Pfam: PF00011; HSP20; 1.  
DR PROSITE: PS01031; HSP20; 1.  
FT NON\_TER 1  
FT NON\_TER 86  
SQ SEQUENCE 86 AA; 9814 MW; 1D6966C08DE5845B CRC64;

Query Match 60.1%; Score 442; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.9e-32;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 FAGLRPTFDTRLMLLEDEMEKGRYEVRAELPGVDPDKVDIMVRDGLTIKARTEQKDF 87  
DB 1 FAGLRPTFDTRLMLLEDEMEKGRYEVRAELPGVDPDKVDIMVRDGLTIKARTEQKDF 60

QY 88 DGRSEFAYGSFVRTVSLPVGADDDI 113  
DB 61 DGRSEFAYGSFVRTVSLPVGADDDI 86

RESULT 8  
Q8KY76\_MYCGN  
ID Q8KY76\_MYCGN PRELIMINARY; PRT; 86 AA.  
AC Q8KY76;  
DT 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Alpha-crystallin protein (Fragment).  
GN Name=acr;  
OS Mycobacterium genavense.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=36812;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=M95-1325R;  
RA Shampula I.C., Baetian I., Fonteyne P.-A., Portaeis F.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)  
family.  
DR EMBL: AF253462; AAM69218.1; -; Genomic\_DNA.  
DR InterPro: IPR002068; Hsp20.  
DR Pfam: PF00011; HSP20; 1.  
DR PROSITE: PS01031; HSP20; 1.  
FT NON\_TER 1  
FT NON\_TER 86  
SQ SEQUENCE 86 AA; 9814 MW; 1D6966C08DE5845B CRC64;

Query Match 60.1%; Score 442; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.9e-32;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 FAGLRPTFDTRLMLLEDEMEKGRYEVRAELPGVDPDKVDIMVRDGLTIKARTEQKDF 87  
DB 1 FAGLRPTFDTRLMLLEDEMEKGRYEVRAELPGVDPDKVDIMVRDGLTIKARTEQKDF 60

QY 88 DGRSEFAYGSFVRTVSLPVGADDDI 113  
DB 61 DGRSEFAYGSFVRTVSLPVGADDDI 86

RESULT 9  
Q8KY72\_MYCCE  
ID Q8KY72\_MYCCE PRELIMINARY; PRT; 86 AA.  
AC Q8KY72;  
DT 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Alpha-crystallin protein (Fragment).  
GN Name=acr;  
OS Mycobacterium celatum.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=28045;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=M95-143;  
RA Shampula I.C., Baetian I., Fonteyne P.-A., Portaeis F.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)  
family.  
DR EMBL: AF253460; AAM69222.1; -; Genomic\_DNA.  
DR InterPro: IPR002068; Hsp20.  
DR Pfam: PF00011; HSP20; 1.  
DR PROSITE: PS01031; HSP20; 1.  
FT NON\_TER 1  
FT NON\_TER 86  
SQ SEQUENCE 86 AA; 9814 MW; 1D6966C08DE5845B CRC64;

Query Match 60.1%; Score 442; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.9e-32;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 FAGLRPTFDTRLMLLEDEMEKGRYEVRAELPGVDPDKVDIMVRDGLTIKARTEQKDF 87  
DB 1 FAGLRPTFDTRLMLLEDEMEKGRYEVRAELPGVDPDKVDIMVRDGLTIKARTEQKDF 60

QY 88 DGRSEFAYGSFVRTVSLPVGADDDI 113  
DB 61 DGRSEFAYGSFVRTVSLPVGADDDI 86

RESULT 10  
Q8KY68\_MYCCH  
ID Q8KY68\_MYCCH PRELIMINARY; PRT; 86 AA.  
AC Q8KY68;  
DT 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Alpha-crystallin protein (Fragment).  
GN Name=acr;  
OS Mycobacterium chelonae.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1774;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=M8223;  
RA Shampula I.C., Baetian I., Fonteyne P.-A., Portaeis F.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)  
family.  
DR EMBL: AF253464; AAM69226.1; -; Genomic\_DNA.  
DR InterPro: IPR002068; Hsp20.  
DR Pfam: PF00011; HSP20; 1.  
DR PROSITE: PS01031; HSP20; 1.  
FT NON\_TER 1  
FT NON\_TER 86  
SQ SEQUENCE 86 AA; 9800 MW; 0FF966C08DE5845B CRC64;

Query Match 59.9%; Score 441; DB 2; Length 86;  
Best Local Similarity 98.8%; Pred. No. 2.4e-32;  
Matches 85; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 28 FAGLRPTFDTRLMLLEDEMEKGRYEVRAELPGVDPDKVDIMVRDGLTIKARTEQKDF 87  
DB 1 FAGLRPTFDTRLMLLEDEMEKGRYEVRAELPGVDPDKVDIMVRDGLTIKARTEQKDF 60

QY 88 DGRSEFAYGSFVRTVSLPVGADDDI 113  
DB 61 DGRSEFAYGSFVRTVSLPVGADDDI 86

```
RESULT 11
Q8KY75 MYCPA PRELIMINARY; PRT; 86 AA.
AC Q8KY75
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN Name:acr;
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium avium complex (MAC).
OX NCBI_TaxID=1770;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=M2672;
RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaeis F.;
RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
CC -1 SIMILARITY: Belongs to the small heat shock protein (HSP20)
family.
DR EMBL, AF253457; AAM69229.1; -; Genomic_DNA.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON_TER 1
FT NON_TER 86
SQ SEQUENCE 86 AA; 9800 MW; 0FF966C08DB5845B CRC64;

Query Match 59.9%; Score 441; DB 2; Length 86;
Best Local Similarity 98.8%; Pred. No. 2.4e-32;
Matches 85; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 28 FAGLRPTFDTRLMLBEDMKEGRYEVRALPGVDPDKVDIMVRDGLTIKAERTEQKF 87
Db 1 FAGLRPTFDTRLMLBEDMKEGRYEVRALPGVDPDKVDIMVRDGLTIKAERTEQKF 60

Qy 88 DGRSEFAYGSFVRTVSLPVGADDDI 113
Db 61 DGRSEFAYGSFVRTVSLPVGADDDV 86

RESULT 12
Q8KY69 MYCRO PRELIMINARY; PRT; 86 AA.
AC Q8KY69
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN Name:acr;
OS Mycobacterium fortuitum.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;
OX NCBI_TaxID=1766;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=M2672;
RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaeis F.;
RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
CC -1 SIMILARITY: Belongs to the small heat shock protein (HSP20)
family.
DR EMBL, AF253453; AAM69225.1; -; Genomic_DNA.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON_TER 1
FT NON_TER 86
SQ SEQUENCE 86 AA; 9770 MW; B76666C0942C5442 CRC64;

Query Match 59.0%; Score 434; DB 2; Length 86;
Best Local Similarity 97.7%; Pred. No. 1e-31;
```

```
Matches 84; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 28 FAGLRPTFDTRLMLBEDMKEGRYEVRALPGVDPDKVDIMVRDGLTIKAERTEQKF 87
Db 1 FAGLRPTFDTRLMLBEDMKEGRYEVRALPGVDPDKVDIMVRDGLTIKAERTEQKF 60

Qy 88 DGRSEFAYGSFVRTVSLPVGADDDI 113
Db 61 DGRSEFAYGSFVRTVSLPVGADDDV 86

RESULT 13
Q8KY67 MYCXE PRELIMINARY; PRT; 86 AA.
AC Q8KY67
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Alpha-crystallin-small heat shock protein (Fragment).
OS Mycobacterium xenopi.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;
OX NCBI_TaxID=1789;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=M2672;
RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaeis F.;
RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
CC -1 SIMILARITY: Belongs to the small heat shock protein (HSP20)
family.
DR EMBL, AF261078; AAM69227.1; -; Genomic_DNA.
DR GO; GO:0006457; P:protein folding; IEA.
DR GO; GO:0006986; P:response to unfolded protein; IEA.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
KM Heat shock.
FT NON_TER 1
FT NON_TER 86
SQ SEQUENCE 86 AA; 9757 MW; BE9BBD0DBACB4E3 CRC64;

Query Match 49.3%; Score 363; DB 2; Length 86;
Best Local Similarity 79.1%; Pred. No. 2.8e-25;
Matches 68; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

Qy 28 FAGLRPTFDTRLMLBEDMKEGRYEVRALPGVDPDKVDIMVRDGLTIKAERTEQKF 87
Db 1 FTGHPITFDRLMLBEDMKEGRYEVRALPGVDPADIDITVRDGLTIKAERTEQKF 60

Qy 88 DGRSEFAYGSFVRTVSLPVGADDDI 113
Db 61 DGRSEFAYGSFVRTVSLPAGANEDDV 86

RESULT 14
Q8KY84 MYCNO PRELIMINARY; PRT; 86 AA.
AC Q8KY84
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN Name:acr;
OS Mycobacterium nonchromogenicum.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;
OX NCBI_TaxID=1782;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATC19530;
RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaeis F.;
RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
CC -1 SIMILARITY: Belongs to the small heat shock protein (HSP20)
family.
```

DR EMBL; AF253445; AAM69207.1; -, Genomic\_DNA.  
 DR InterPro; IPR02068; HSP20.  
 DR Pfam; PF00011; HSP20; 1.  
 DR PROSITE; PS01031; HSP20; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 86 AA; 9722 MW; BAF085B56BF1DBF3 CRC64;

Query Match 46.6%; Score 343; DB 2; Length 86;  
 Best Local Similarity 74.4%; Pred. No. 1.9e-23;  
 Matches 64; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

QY 28 FAGLRPTFDTRLMLEDEMKEGREYRAELPGVDPDKVDIMVRDQGLTIKAERTQKDF 87  
 Db 1 FAGLRPAFDSRLMLEDEMKGDRYELRAELPGVDPKEDIETVLDGQLTIKAERSKKEF 60

QY 88 DGRSEFAYGSFVRTVSLPVGADDEDDI 113  
 Db 61 NGRSEFCYGSFRTVPLPAGVTEDGI 86

## RESULT 15

Q8KY71 MYCGS  
 ID Q8KY71\_MYCGS PRELIMINARY; PRT; 86 AA.  
 AC Q8KY71;  
 DT 01-OCT-2002 (TReMBLrel. 22, Created)  
 DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Alpha-crystallin protein (Fragment).  
 GN Name=acr;  
 OS Mycobacterium gastri.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_Taxid=1777;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=M97-938;  
 RA Shampita I.C., Bastian I., Fonteyne P.-A., Portaeals F.;  
 RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)  
 family.  
 CC EMBL; AF253461; AAM69223.1; -, Genomic\_DNA.  
 DR InterPro; IPR02068; HSP20.  
 DR Pfam; PF00011; HSP20; 1.  
 DR PROSITE; PS01031; HSP20; 1.  
 FT NON\_TER 1  
 FT NON\_TER 86  
 SQ SEQUENCE 86 AA; 9722 MW; BAF085B56BF1DBF3 CRC64;

Query Match 46.6%; Score 343; DB 2; Length 86;  
 Best Local Similarity 74.4%; Pred. No. 1.9e-23;  
 Matches 64; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

QY 28 FAGLRPTFDTRLMLEDEMKEGREYRAELPGVDPDKVDIMVRDQGLTIKAERTQKDF 87  
 Db 1 FAGLRPAFDSRLMLEDEMKGDRYELRAELPGVDPKEDIETVLDGQLTIKAERSKKEF 60

QY 88 DGRSEFAYGSFVRTVSLPVGADDEDDI 113  
 Db 61 NGRSEFCYGSFRTVPLPAGVTEDGI 86

Search completed: March 23, 2006, 05:24:22  
 Job time : 38.8298 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using BW model

Run on: March 23, 2006, 05:24:53 ; Search time 9.82128 Seconds

(without alignments)  
1212.194 Million cell updates/sec

Title: US-10-617-038-24

Perfect score: 736  
Sequence: 1 MATTLPVQRHRSIFPERSF.....SVAVSEKPTKHIQIRSTN 144

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/1/1aa/5\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/6\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/H\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/PCUTS\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/RE\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	736	100.0	144	2	US-09-462-556-2
2	736	100.0	331	2	US-09-223-040-4
3	736	100.0	331	2	US-09-287-849-4
4	736	100.0	368	2	US-09-287-849-24
5	736	100.0	433	2	US-09-287-849-14
6	736	100.0	856	2	US-09-287-849-12
7	718	97.6	143	2	US-09-461-774-8
8	245	33.3	48	2	US-09-157-689-18
9	245	33.3	48	2	US-08-447-398-18
10	245	33.3	48	2	US-09-953-510-18
11	146.5	17.9	151	2	US-09-902-540-13672
12	132	17.9	139	2	US-09-107-532A-4324
13	116	15.8	131	2	US-09-902-540-11500
14	100.5	13.7	250	2	US-09-949-016-8090
15	94	12.8	156	2	US-09-543-681A-7799
16	93.5	12.7	167	2	US-09-835-909A-2
17	84.5	11.5	759	2	US-09-793-024-76
18	83.5	11.3	177	2	US-10-105-427-2
19	82.5	11.2	274	2	US-09-107-532A-4705
20	81	11.0	460	2	US-09-199-637A-132
21	79	10.7	175	2	US-09-513-999C-7967
22	79	10.7	183	2	US-09-949-016-8416
23	79	10.7	250	2	US-09-513-999C-7968
24	79	10.7	918	2	US-09-198-452A-1072
25	79	10.7	918	2	US-09-438-185A-999
26	78.5	10.7	475	2	US-09-252-991A-79494
27	75.5	10.3	199	1	US-08-900-407-4

28	75.5	10.3	240	2	US-09-949-016-10863	Sequence 10863, A
29	75.5	10.3	459	2	US-09-513-783A-170	Sequence 170, App
30	75.5	10.3	459	2	US-10-100-957A-170	Sequence 170, App
31	74.5	10.1	168	2	US-09-949-016-8494	Sequence 8494, Ap
32	73	9.9	291	2	US-09-902-540-12550	Sequence 12550, A
33	72.5	9.9	545	2	US-09-902-540-10533	Sequence 10533, A
34	72	9.8	188	2	US-09-902-540-10533	Sequence 9929, Ap
35	72	9.8	332	2	US-09-605-703B-2268	Sequence 2268, Ap
36	71.5	9.7	193	2	US-08-900-407-3	Sequence 3, Appli
37	71.5	9.7	311	2	US-09-543-681A-5563	Sequence 5563, Ap
38	71.5	9.7	693	2	US-09-376-343-2	Sequence 2, Appli
39	71	9.6	497	2	US-09-328-352-7286	Sequence 7286, Ap
40	70.5	9.6	923	2	US-09-328-352-4371	Sequence 4371, Ap
41	70.5	9.6	1132	2	US-09-248-796A-15026	Sequence 15026, A
42	68.5	9.3	167	2	US-09-489-039A-13896	Sequence 13896, A
43	68.5	9.3	486	2	US-09-252-991A-24336	Sequence 24336, A
44	68.5	9.3	594	2	US-09-252-991A-18099	Sequence 18099, A
45	68	9.2	248	2	US-09-489-039A-11767	Sequence 11767, A

## ALIGNMENTS

```

RESULT 1
US-09-462-556-2
; Sequence 2, Application US/09462556
; Patent No. 6403100
; GENERAL INFORMATION:
; APPLICANT: Barry III, Clifton E.
; APPLICANT: Yuan, Ying
; APPLICANT: Crane, Deborah D.
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by The Secretary of the
; APPLICANT: Department of Health and Human Services
; TITLE OF INVENTION: Method of Attenuating Pathogenic Mycobacteria and
; TITLE OF INVENTION: Strains of Mycobacteria So Attenuated
; FILE REFERENCE: 015280-333100US
; CURRENT APPLICATION NUMBER: US/09/462,556
; CURRENT FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: US 60/052,199
; PRIOR FILING DATE: 1997-07-10
; PRIOR APPLICATION NUMBER: WO PCT/US98/14227
; PRIOR FILING DATE: 1998-07-09
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 144
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: alpha-crystallin heat shock protein
US-09-462-556-2

Query Match      100.0%; Score 736; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 9.8e-81;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MATTLPVQRHRSIFPERSLFAAFSPAGIRPFDRIMLDEBMEKRGREYVAELPGV 60
      |||||||
DB      1 MATTLPVQRHRSIFPERSLFAAFSPAGIRPFDRIMLDEBMEKRGREYVAELPGV 60
      |||||||

QY      61 DPDDVDIMVWDGQLTIAKRTKQDFDGRSEFVYGSFVRTVSLPVGADEDDIKATYDKG 120
      |||||||
DB      61 DPDDVDIMVWDGQLTIAKRTKQDFDGRSEFVYGSFVRTVSLPVGADEDDIKATYDKG 120
      |||||||

QY      121 ILTVSVAVSEKPTKHIQIRSTN 144
      |||||||
DB      121 ILTVSVAVSEKPTKHIQIRSTN 144
      |||||||

RESULT 2
US-09-223-040-4
; Sequence 4, Application US/09223040

```

```
Patent No. 6544522
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009010US
CURRENT APPLICATION NUMBER: US/09/223,040
CURRENT FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 331
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:tr1-fusion
US-09-223-040-4
```

```
Query Match          100.0%; Score 736; DB 2; Length 331;
Best Local Similarity 100.0%; Pred. No. 3,4e-80;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MATLPGVQRHPRSLFPEFSELFAAPSPAGLRPTFTDRLMLRLEDEMKEGRYEVRAELPGV 60
    |||||||
DB 8 MATLPGVQRHPRSLFPEFSELFAAPSPAGLRPTFTDRLMLRLEDEMKEGRYEVRAELPGV 67
    |||||||
QY 61 DPKDQDVIMVRDGLTIKARTEQKDFGRSEFAYGSFVRTVSLPVGADEDDIKATYDKG 120
    |||||||
DB 68 DPKDQDVIMVRDGLTIKARTEQKDFGRSEFAYGSFVRTVSLPVGADEDDIKATYDKG 127
    |||||||
QY 121 ILTVSAVASEGKPTKEKHQIRSTN 144
    |||||||
DB 128 ILTVSAVASEGKPTKEKHQIRSTN 151
    |||||||
```

```
RESULT 3
US-09-287-849-4
Sequence 4, Application US/09287849
Patent No. 6627198
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/09/287,849
CURRENT FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 331
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:tr1-fusion
US-09-287-849-4
```

```
Query Match          100.0%; Score 736; DB 2; Length 331;
Best Local Similarity 100.0%; Pred. No. 3,4e-80;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MATLPGVQRHPRSLFPEFSELFAAPSPAGLRPTFTDRLMLRLEDEMKEGRYEVRAELPGV 60
    |||||||
DB 8 MATLPGVQRHPRSLFPEFSELFAAPSPAGLRPTFTDRLMLRLEDEMKEGRYEVRAELPGV 67
    |||||||
QY 61 DPKDQDVIMVRDGLTIKARTEQKDFGRSEFAYGSFVRTVSLPVGADEDDIKATYDKG 120
    |||||||
DB 68 DPKDQDVIMVRDGLTIKARTEQKDFGRSEFAYGSFVRTVSLPVGADEDDIKATYDKG 127
    |||||||
QY 121 ILTVSAVASEGKPTKEKHQIRSTN 144
    |||||||
DB 128 ILTVSAVASEGKPTKEKHQIRSTN 151
    |||||||
```

```
RESULT 4
US-09-287-849-24
Sequence 24, Application US/09287849
Patent No. 6627198
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/09/287,849
CURRENT FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 24
LENGTH: 368
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:tr1-fusion
US-09-287-849-24
```

```
Query Match          100.0%; Score 736; DB 2; Length 368;
Best Local Similarity 100.0%; Pred. No. 4e-80;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MATLPGVQRHPRSLFPEFSELFAAPSPAGLRPTFTDRLMLRLEDEMKEGRYEVRAELPGV 60
    |||||||
DB 9 MATLPGVQRHPRSLFPEFSELFAAPSPAGLRPTFTDRLMLRLEDEMKEGRYEVRAELPGV 68
    |||||||
QY 61 DPKDQDVIMVRDGLTIKARTEQKDFGRSEFAYGSFVRTVSLPVGADEDDIKATYDKG 120
    |||||||
DB 69 DPKDQDVIMVRDGLTIKARTEQKDFGRSEFAYGSFVRTVSLPVGADEDDIKATYDKG 128
    |||||||
QY 121 ILTVSAVASEGKPTKEKHQIRSTN 144
    |||||||
DB 129 ILTVSAVASEGKPTKEKHQIRSTN 152
    |||||||
```

```
RESULT 5
US-09-287-849-14
```

```
; Sequence 14, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neco, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-14

Query Match          100.0%; Score 736; DB 2; Length 433;
Best Local Similarity 100.0%; Pred. No. 5.2e-80;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATLLPVQRHPRSLFPPFSELPFAAPSPFAGLRPTFDTRLMLRLEDEMKRGYEVRAELPGV 60
Db 9 MATLLPVQRHPRSLFPPFSELPFAAPSPFAGLRPTFDTRLMLRLEDEMKRGYEVRAELPGV 68

Qy 61 DPDKVDIMVRDGLTTIKARTEQKDPGRSEFAVGSFVRTVSLPVGADDDIKATYDKG 120
Db 69 DPDKVDIMVRDGLTTIKARTEQKDPGRSEFAVGSFVRTVSLPVGADDDIKATYDKG 128

Qy 121 LITVSAVSEKPTKHIQIRSTN 144
Db 129 LITVSAVSEKPTKHIQIRSTN 152

RESULT 6
US-09-287-849-12
; Sequence 12, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neco, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
```

```
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:penta-fusion
US-09-287-849-12

Query Match          100.0%; Score 736; DB 2; Length 856;
Best Local Similarity 100.0%; Pred. No. 1.4e-79;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATLLPVQRHPRSLFPPFSELPFAAPSPFAGLRPTFDTRLMLRLEDEMKRGYEVRAELPGV 60
Db 9 MATLLPVQRHPRSLFPPFSELPFAAPSPFAGLRPTFDTRLMLRLEDEMKRGYEVRAELPGV 68

Qy 61 DPDKVDIMVRDGLTTIKARTEQKDPGRSEFAVGSFVRTVSLPVGADDDIKATYDKG 120
Db 69 DPDKVDIMVRDGLTTIKARTEQKDPGRSEFAVGSFVRTVSLPVGADDDIKATYDKG 128

Qy 121 LITVSAVSEKPTKHIQIRSTN 144
Db 129 LITVSAVSEKPTKHIQIRSTN 152

RESULT 7
US-09-461-774-8
; Sequence 8, Application US/09461774
; Patent No. 6887481
; GENERAL INFORMATION:
; APPLICANT: CHAN, Lily
; APPLICANT: CHUNG, Maxey Ching Ming
; APPLICANT: LIM, Renee Lay Hong
; TITLE OF INVENTION: Bacterial-derived molecules and therapeutic and
; TITLE OF INVENTION: diagnostic uses therefor
; FILE REFERENCE: 1781-0180P
; CURRENT APPLICATION NUMBER: US/09/461,774
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-461-774-8

Query Match          97.6%; Score 718; DB 2; Length 143;
Best Local Similarity 98.6%; Pred. No. 1.4e-78;
Matches 141; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ATLLPVQRHPRSLFPPFSELPFAAPSPFAGLRPTFDTRLMLRLEDEMKRGYEVRAELPGV 61
Db 1 ATLLPVQRHPRSLFPPFSELPFAAPSPFAGLRPTFDTRLMLRLEDEMKRGYEVRAELPGV 60

Qy 62 PDKVDIMVRDGLTTIKARTEQKDPGRSEFAVGSFVRTVSLPVGADDDIKATYDKGI 121
Db 61 PDKVDIMVRDGLTTIKARTEQKDPGRSEFAVGSFVRTVSLPVGADDDIKATYDKGI 120

Qy 122 LITVSAVSEKPTKHIQIRSTN 144
Db 121 LITVSAVSEKPTKHIQIRSTN 143

RESULT 8
US-09-157-689-18
; Sequence 18, Application US/09157689
; Patent No. 6599510
```



```

: GENERAL INFORMATION:
: APPLICANT: Horwitz, Marcus A.
: APPLICANT: Harth, Gunter
: TITLE OF INVENTION: Abundant Extracellular Products and
: TITLE OF INVENTION: Methods for Their Production and Use
: NUMBER OF SEQUENCES: 95
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kurt A. Maclean
: STREET: 2029 Century Park East, Suite 3800
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90067
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/157,689
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/568,357
: FILING DATE: 06-DEC-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/551,149
: FILING DATE: 31-OCT-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/447,398
: FILING DATE: 23-MAY-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/289,667
: FILING DATE: 12-AUG-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/156,358
: FILING DATE: 23-NOV-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Maclean, Kurt A.
: REGISTRATION NUMBER: 31,118
: REFERENCE/DOCKET NUMBER: 118-119
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (310) 788-5000
: TELEFAX: (310) 277-1297
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 48 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: N-terminal
: ORIGINAL SOURCE:
: ORGANISM: Mycobacterium tuberculosis
: STRAIN: Erdman
: US-09-157-689-18

Query Match      33.3%; Score 245; DB 2; Length 48;
Best Local Similarity 100.0%; Pred. No. 2.9e-22;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      35  FDTRLMLDEDMKEGRYVRAELPGVDPDKVDIMVRDGLTIKAERT 82
Db      1  FDTRLMLDEDMKEGRYVRAELPGVDPDKVDIMVRDGLTIKAERT 48

RESULT 9
US-08-447-398-18
: Sequence 18, Application US/08447398
: Patent No. 6761894
: GENERAL INFORMATION:

```

```

: APPLICANT: Horwitz, Marcus A.
: TITLE OF INVENTION: Abundant Extracellular Products and
: TITLE OF INVENTION: Methods for Their Production and Use
: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kurt A. Maclean
: STREET: 2029 Century Park East, Suite 3800
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90067
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/447,398
: FILING DATE: 23-MAY-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/289,667
: FILING DATE: 12-AUG-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/156,358
: FILING DATE: 23-NOV-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Maclean, Kurt A.
: REGISTRATION NUMBER: 31,118
: REFERENCE/DOCKET NUMBER: 111-161
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (310) 788-5000
: TELEFAX: (310) 277-1297
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 48 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: N-terminal
: ORIGINAL SOURCE:
: ORGANISM: Mycobacterium tuberculosis
: STRAIN: Erdman
: US-08-447-398-18

Query Match      33.3%; Score 245; DB 2; Length 48;
Best Local Similarity 100.0%; Pred. No. 2.9e-22;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      35  FDTRLMLDEDMKEGRYVRAELPGVDPDKVDIMVRDGLTIKAERT 82
Db      1  FDTRLMLDEDMKEGRYVRAELPGVDPDKVDIMVRDGLTIKAERT 48

RESULT 10
US-09-953-510-18
: Sequence 18, Application US/09953510
: Patent No. 6818223
: GENERAL INFORMATION:
: APPLICANT: Horwitz, Marcus A.
: TITLE OF INVENTION: Abundant Extracellular
: Products and Methods for Their Production and Use
: NUMBER OF SEQUENCES: 91
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kurt A. Maclean
: STREET: 2029 Century Park East, Suite 3800
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90067

```

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0,  
Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/953,510  
FILING DATE: 14-Sep-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/447,398  
FILING DATE: 23-MAY-1995  
APPLICATION NUMBER: US 08/289,667  
FILING DATE: 12-AUG-1994  
APPLICATION NUMBER: US 08/156,358  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: MacLean, Kurt A.  
REGISTRATION NUMBER: 31,118  
REFERENCE/DOCKET NUMBER: 112-272  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (310) 788-5000  
TELEFAX: (310) 277-1297  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Mycobacterium tuberculosis  
STRAIN: Erdman  
SEQUENCE DESCRIPTION: SEQ ID NO: 18:  
US-09-953-510-18  
Query Match 33.3%; Score 245; DB 2; Length 48;  
Best Local Similarity 100.0%; Pred. No. 2,9e-22;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 35 FDRFLMLEDEMKGRYEVRAELPGVDPDKVDIMVRDGLTIKAERT 82  
Db 1 FDRFLMLEDEMKGRYEVRAELPGVDPDKVDIMVRDGLTIKAERT 48  
RESULT 11  
US-09-902-540-13672  
Sequence 13672, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(115849)B  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/4217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 13672  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Myxococcus xanthus  
US-09-902-540-13672  
Query Match 19.9%; Score 146.5; DB 2; Length 151;  
Best Local Similarity 33.1%; Pred. No. 1.2e-09;

Matches 39; Conservative 29; Mismatches 31; Indels 19; Gaps 6;  
Qy 30 GLRPTFDRFLMLEDEMKGR--YEVRAELPGVDPDKVDIMVRDGLTIKAERTQKDF 87  
Db 35 GFLPDPF-----EVKRTQDAFIRKADVPGE-EDLEITLAEKRLTISGRREBRD 84  
Qy 88 DGRSEFA---YGSFVRYTSLPVGADEDDIKATYDKILTVSV-ANSEKPTKHIOI 140  
Db 85 EGRRYVAYERNYGSFSRTFTLPGRVANADVQDFKGLNVRIPKXSEOP--KRIRV 140  
RESULT 12  
US-09-107-532A-4324  
Sequence 4324, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucetere-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Walham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariadello, Pamela Deneka  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781) 893-5007  
TELEFAX: (781) 893-8277  
INFORMATION FOR SEQ ID NO: 4324:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 139 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...139  
SEQUENCE DESCRIPTION: SEQ ID NO: 4324:  
US-09-107-532A-4324  
Query Match 17.9%; Score 132; DB 2; Length 139;  
Best Local Similarity 33.1%; Pred. No. 6.1e-08;  
Matches 46; Conservative 26; Mismatches 47; Indels 20; Gaps 10;  
Qy 12 RSLPFRSEFLA-AFSPFAGLR--PTFDRFLMLEDEMKGRYEVRAELPGVDPDKVDI 68  
Db 10 RDMFPDNDVFSRAFNDFLVSSYPKVD-----LVNEXE--YKLTADMFGCDKE-DTV 61  
Qy 69 MVRDGLTIYA---ERTEQDFDGC--RSEFAYGSFVRYTSLPVGADEDDIKATYDKIL 122  
Db 62 EYSDNTLTISANSHSHTEDEK-DGNVYRKERHSHSVYSKRSFYL-P-NVDEKITGTFKXGVL 119

```
QY      123 TVSVAVSECKPTE-KHIQI 140
        ::::|:::
DB      120 KLVLPKTAQPKETKIEL 138
```

```

RESULT 13
US-09-902-540-11500
: Sequence 11500, Application US/09902540
: Patent No. 6833447
: GENERAL INFORMATION:
: APPLICANT: Goldman, Barry S.
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: APPLICANT: Wiegand, Roger C.
: TITLE OF INVENTION: MYXOCOCCUS xanthus Genome Sequences and Uses Thereof
: FILE REFERENCE: 38-10(1)58491B
: CURRENT APPLICATION NUMBER: US/09/902,540
: CURRENT FILING DATE: 2001-07-10
: PRIOR APPLICATION NUMBER: 60/217,883
: PRIOR FILING DATE: 2000-07-10
: NUMBER OF SEQ ID NOS: 16825
: SEQ ID NO 11500
: LENGTH: 131
: TYPE: prt
: ORGANISM: Myxococcus xanthus
: US-09-902-540-11500

```

```

RESULT 14
US-09-949-016-8090
: Sequence 8090, Application US/09949016
: Patent No. 681235
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: CL0010107
: CURRENT APPLICATION NUMBER: US/09/949, 016
: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 8090
: LENGTH: 250
: TYPE: PRT
: ORGANISM: Human
: US-09-949-016-8090

```

```

OY          119 KG 120
DB          192 LG 193

RESULT 15
US-09-543-681A-7799
; Sequence 7799, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543, 681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128, 706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7799
; LENGTH: 156
; TYPE: prt
; ORGANISM: Proteus mirabilis
US-09-543-681A-7799

```

Search completed: March 23, 2006, 05:30:14  
Job time : 10.9213 secs



```

1 CURRENT APPLICATION NUMBER: US/10/332,512A
2
3 CURRENT FILING DATE: 2003-01-10
4
5 PRIOR APPLICATION NUMBER: PCT/US01/21717
6
7 PRIOR FILING DATE: 2001-07-10
8
9 PRIOR APPLICATION NUMBER: US 60/217,646
10
11 PRIOR FILING DATE: 2000-07-10
12
13 NUMBER OF SEQ ID NOS: 31
14
15 SOFTWARE: PatentIn version 3.1
16
17 SEQ ID NO 23
18
19 LENGTH: 144
20
21 TYPE: PR1
22
23 ORGANISM: Mycobacterium tuberculosis
24
25 US-10-332-512A-23

```

Query Match	100.0%	Score 736;	DB 4;	Length 144;
Best Local Similarity	100.0%;	Pred. No. 1,1e-72;		
Matches 144;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
QY	1	MATLLPVGRHRSLEPFESSELPFAAPPSFAGLRPTFDTRLMRLEDEMKGRGYVRALPGV	60	
Db	1	MATLLPVGRHRSLEPFESSELPFAAPPSFAGLRPTFDTRLMRLEDEMKGRGYVRALPGV	60	
QY	61	DPDQVDIMARDGQTTKAERTEQKDPGGRSEFAFGSTVRYTSLPVGADDEDDIKATYDKG	120	
Db	61	DPDQVDIMARDGQTTKAERTEQKDPGGRSEFAFGSTVRYTSLPVGADDEDDIKATYDKG	120	
QY	121	ILTVSVAVSEKGPTEKHQIRSTN	144	
Db	121	ILTVSVAVSEKGPTEKHQIRSTN	144	

```

RESULT 3
US-09-287-849-4
; Sequence 4, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 331
; TYPE: PRN
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; US-09-287-849-4

```

Query Match	100.0%	Score 736:	DB 3:	Length 331:
Best Local Similarity	100.0%	Pred. No. 3.5e-72:		
Matches 144:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:
QY	1	MATLTPQQRHRSLPFPERSELPAAFPSPAGLRPTFTDTLMLLEDEMKRGREYVRAELPGV	60	

D6	8	MATTLIPQIHPSLSLPEFBSFSELEAAPPSPFAGLPTPTDITLMLBEMKGRYEVAELPGV	67
QY	61	DPDKVDIYWRDQOLITKAERTEQDPDPSRSEFAAGSEVFTYSLPEVGADDEDIKATYDKG	120
Db	68	DPDKVDIYWRDQOLITKAERTEQDPDPSRSEFAAGSEVFTYSLPEVGADDEDIKATYDKG	122
QY	121	ILTVSVANSEGRKTEKHQIRSTN	144
Db	128	ILTVSVANSEGRKTEKHQIRSTN	151

```

1 RESULT 4
2 US-10-359-460--4
3 ; Sequence 4, Application US/10359460
4 ; Publication No. US20030147911A1
5 ; GENERAL INFORMATION:
6 ; APPLICANT: Reed, Steven G.
7 ; APPLICANT: Skeiky, Yasir A.W.
8 ; APPLICANT: Dillon, Davin C.
9 ; APPLICANT: Alderson, Mark
10 ; APPLICANT: Campos-Neto, Antonio
11 ; APPLICANT: Corixa Corporation
12 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
13 ; TITLE OF INVENTION: and Their Uses
14 ; FILE REFERENCE: 014058-009020US
15 ; CURRENT APPLICATION NUMBER: US/10/359,460
16 ; CURRENT FILING DATE: 2003-02-05
17 ; PRIOR APPLICATION NUMBER: US/09/287,849
18 ; PRIOR FILING DATE: 1999-04-07
19 ; PRIOR APPLICATION NUMBER: US 08/818,112
20 ; PRIOR FILING DATE: 1997-03-13
21 ; PRIOR APPLICATION NUMBER: US 08/942,578
22 ; PRIOR FILING DATE: 1997-10-01
23 ; PRIOR APPLICATION NUMBER: US 09/025,197
24 ; PRIOR FILING DATE: 1998-02-18
25 ; PRIOR APPLICATION NUMBER: US 09/056,556
26 ; PRIOR FILING DATE: 1998-04-07
27 ; PRIOR APPLICATION NUMBER: US 09/223,040
28 ; PRIOR FILING DATE: 1998-12-30
29 ; NUMBER OF SEQ ID NOS: 46
30 ; SOFTWARE: PatentIn Ver. 2.1
31 ; SEQ ID NO 4
32 ; LENGTH: 331
33 ; TYPE: PR1
34 ; ORGANISM: Artificial Sequence
35 ; FEATURE:
36 ; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
37 ; US-10-359-460--4

```

```
Query Match      100.0%; Score 736; DB 4; Length 331;
Beet Local Similarity 100.0%; Pred. NO. 3.5e-72;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 MATTLPVQRHPRSLFPFSSELFAAFPSFAGLRPTFDTRLMLRLEDMKRGYVRAELPGV 60
        |||
Db      8 MATLTPVQRHPRSLFPFSSELFAAFPSFAGLRPTFDTRLMLRLEDMKRGYVRAELPGV 67

Qy      61 DPDVDVIDMRDGLTTKAERTEKDDPGRSEFYGSFVRVLSLVGADDEDIDATYDKG 120
        |||
Db      68 DDPDVDVIDMRDGLTTKAERTEKDDPGRSEFYGSFVRVLSLVGADDEDIDATYDKG 127

Qy      121 ILTVSVAVSEGGKPTTEKHIIQIRSTN 144
        |||
Db      128 ILTVSVAVSEGGKPTTEKHIIQIRSTN 151

RESULT 5
US-10-359-459-4
; Sequence 4, Application US/10359459
; Publication No. US20040013677A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yashir
; APPLICANT: Alderson, Mark
```

```

1  APPLICANT: Campos-Neto, Antonio
2  TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
3  TITLE OF INVENTION: and Their Uses
4  FILE REFERENCE: 014058-009020US
5  CURRENT APPLICATION NUMBER: US/09/287,849
6  CURRENT FILING DATE: 1999-04-07
7  PRIOR APPLICATION NUMBER: US 08/819,112
8  PRIOR FILING DATE: 1997-03-13
9  PRIOR APPLICATION NUMBER: US 08/942,578
10 PRIOR FILING DATE: 1997-10-01
11 PRIOR APPLICATION NUMBER: US 09/025,157
12 PRIOR FILING DATE: 1998-02-18
13 PRIOR APPLICATION NUMBER: US 09/056,556
14 PRIOR FILING DATE: 1998-04-07
15 PRIOR APPLICATION NUMBER: US 09/223,040
16 PRIOR FILING DATE: 1998-12-30
17 NUMBER OF SEQ ID NOS: 46
18 SOFTWARE: PatentIn Ver. 2.1
19 SEQ ID NO 24
20 LENGTH: 368
21 TYPE: PRT
22 ORGANISM: Artificial Sequence
23 FEATURE:
24 OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
25 US-09-287-849-24

```

Qy	1	MATLLPVQRHRSIFPFSSELFAA	FPFAGLRPFDTRLMRLEDEMEGREYVAAL	PGV	60
Db	9	MATLLPVQRHRSIFPFSSELFAA	FPFAGLRPFDTRLMRLEDEMEGREYVAAL	PGV	68
Qy	61	DPDKDVIIMRDGLTIKARTKEDKD	DGRSEFAYGSFVRVSLPVGADDEDI	KATYDKG	120
Db	69	DPDKDVIIMRDGLTIKARTKEDKD	DGRSEFAYGSFVRVSLPVGADDEDI	KATYDKG	128
Qy	121	ILITVAVSESGKPTKHIQIRSTN			144
Db	129	ILITVAVSESGKPTKHIQIRSTN			152

RESULT 8  
US-09-287-849-14

```
Sequence 14, Application US/09287849
Patent No. US2002009459A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/09/287, 849
PRIOR FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 14
LENGTH: 433
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-14

Query Match      100.0%; Score 736; DB 3; Length 433;
Best Local Similarity 100.0%; Pred. No.5e-72;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATTLPVQRHPSLFPFSESLFAAFSPFAGLRPTFTRLRLDEMEKGRYEVRAELPGV 60
DB 9 MATTLPVQRHPSLFPFSESLFAAFSPFAGLRPTFTRLRLDEMEKGRYEVRAELPGV 68
QY 61 DPKDQVDIMVRDQGLTIKARTEQKDFGRSEFAYGSFVTVSLPVGADEDDIKATYDKG 120
DB 69 DPKDQVDIMVRDQGLTIKARTEQKDFGRSEFAYGSFVTVSLPVGADEDDIKATYDKG 128
QY 121 ILTVSAVSEKGPTEKHIIQIRSTN 144
DB 129 ILTVSAVSEKGPTEKHIIQIRSTN 152

RESULT 9
US-10-359-460-14
Sequence 14, Application US/10359460
Publication No. US2003014791A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/10/359, 460
CURRENT FILING DATE: 2003-02-05
PRIOR APPLICATION NUMBER: US/09/287, 849
PRIOR FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
```

```
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 14
LENGTH: 433
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-10-359-460-14

Query Match      100.0%; Score 736; DB 4; Length 433;
Best Local Similarity 100.0%; Pred. No.5e-72;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATTLPVQRHPSLFPFSESLFAAFSPFAGLRPTFTRLRLDEMEKGRYEVRAELPGV 60
DB 9 MATTLPVQRHPSLFPFSESLFAAFSPFAGLRPTFTRLRLDEMEKGRYEVRAELPGV 68
QY 61 DPKDQVDIMVRDQGLTIKARTEQKDFGRSEFAYGSFVTVSLPVGADEDDIKATYDKG 120
DB 69 DPKDQVDIMVRDQGLTIKARTEQKDFGRSEFAYGSFVTVSLPVGADEDDIKATYDKG 128
QY 121 ILTVSAVSEKGPTEKHIIQIRSTN 144
DB 129 ILTVSAVSEKGPTEKHIIQIRSTN 152

RESULT 10
US-09-287-849-12
Sequence 12, Application US/09287849
Patent No. US2002009459A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/09/287, 849
CURRENT FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 12
LENGTH: 856
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:penta-fusion
US-09-287-849-12

Query Match      100.0%; Score 736; DB 3; Length 856;
Best Local Similarity 100.0%; Pred. No.1.3e-71;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	1	MATLTPQRRPRSLPFPESESLFAAPSPAGLPTPTDRLMRLEDEMKGREVAALPGV	60
Db	9	MATLTPQRRPRSLPFPESESLFAAPSPAGLPTPTDRLMRLEDEMKGREVAALPGV	68
Qy	61	DPDKVDIMWRDQLTIKARTEQKQPDFRSSEFAYGSPVRYTVSLPVGADEDIDIKATYDKG	120
Db	69	DPDKVDIMWRDQLTIKARTEQKQPDFRSSEFAYGSPVRYTVSLPVGADEDIDIKATYDKG	128
Qy	121	ILTVSAVSEKGPTEKHQIQRSTN	144
Db	129	ILTVSAVSEKGPTEKHQIQRSTN	152

```

RESULT 11
US-10-359-460-12
; Sequence 12. Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 856
; TYPE: PRM
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:penta-fusion
; US-10-359-460-12

```

```

Query Match Similarity 100.0%; Score 736; DB 4; length 856;
Best Local Similarity 100.0%; Pred. No. 1.3e-71;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY      1  MATTLPVGRHPRSLPFESELPFAFPSPFAGLRPTFDYTRIMRLADEMKEGRYEVRAELPGV 60
Db      9  MATTLPVGRHPRSLPFESELPFAFPSPFAGLRPTDTRIMRLADEMKEGRYEVRAELPGV 68

QY      61  DDDKDVIDIMVRDGLTIKAERTEQDPOGRSEFAYGSPFRITSLPVGADEDDIKATYDKG 120
Db      69  DDDKDVIDIMVRDGLTIKAERTEQDPOGRSEFAYGSPFRITSLPVGADEDDIKATYDKG 128

QY      121  ILTVSVAVSEGGKPTKHIQIRSTN 144
Db      129  ILTVSVAVSEGGKPTKHIQIRSTN 152

RESULT 12
US-10-369-983-13
; Sequence 13, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:

```

```

APPLICANT: Skeiky, Yasir
APPLICANT: Guderian, Jeff
APPLICANT: Reed, Steven
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
FILE REFERENCE: 014058-009081US
CURRENT APPLICATION NUMBER: US/10/369,983
CURRENT FILING DATE: 2003-02-18
PRIORITY APPLICATION NUMBER: US 60/357,351
PRIORITY FILING DATE: 2002-02-15
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 13
LENGTH: 875
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: fusion protein
OTHER INFORMATION: MTB89F (MTB72F-Erd14)
US-10-369-983-13

```

Query Match Similarity	100.0%	Score 736	DB 4	Length 875
Best Local Similarity	100.0%	Pred. No. 1,3e-71		
Matches 144	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MATTLPVORHRSLSFPEPSELFAAFPSFAGLRPTFDTRLMLRLEDEMKRGREYVRAELPGV	60	
DB	732	MATTLPVORHRSLSFPEPSELFAAFPSFAGLRPTFDTRLMLRLEDEMKRGREYVRAELPGV	791	
QY	61	DDPKVDIMVWDGQLTKAERTBOKDPDGRSEFPAYGSFVTVSLPVGADDEDDIKATYDKG	120	
DB	792	DDPKVDIMVWDGQLTKAERTBOKDPDGRSEFPAYGSFVTVSLPVGADDEDDIKATYDKG	851	
QY	121	ILTVSVAVSEKGPTEKHQIRSTN	144	
DB	852	ILTVSVAVSEKGPTEKHQIRSTN	875	

RESULT 13  
 US-09-953-510-18  
 Sequence 18, Application US/09953510  
 Patent No. US20020131975A1  
 GENERAL INFORMATION:  
 APPLICANT: Horwitz, Marcus A.  
 TITLE OF INVENTION: Abundant Extracellular  
 Products and Methods for Their Production and Use  
 NUMBER OF SEQUENCES: 91  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Kurt A. Maclean  
 STREET: 2029 Century Park East, Suite 3800  
 CITY: Los Angeles  
 STATE: California  
 COUNTRY: U.S.A.  
 ZIP: 90067  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0,  
 Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/953,510  
 FILING DATE: 14-Sep-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/447,398  
 FILING DATE: 23-MAY-1995  
 APPLICATION NUMBER: US 08/289,667  
 FILING DATE: 12-AUG-1994  
 APPLICATION NUMBER: US 08/156,358  
 FILING DATE: 23-NOV-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Maclean, Kurt A.



REGISTRATION NUMBER: 31,118  
REFERENCE/DOCKET NUMBER: 112-272  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (310) 788-5000  
TELEFAX: (310) 277-1297  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Mycobacterium tuberculosis  
STRAIN: Erdman  
SEQUENCE DESCRIPTION: SEQ ID NO: 18:  
US-09-953-510-18

Query Match 33.3%; Score 245; DB 3; Length 48;  
Best Local Similarity 100.0%; Pred. No. 2.5e-19;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 35 FDTLRMLRLEDEMKGRYEVRAELPGVDPDQVDIMVRDGLTIKAERT 82  
Db 1 FDTLRMLRLEDEMKGRYEVRAELPGVDPDQVDIMVRDGLTIKAERT 48

RESULT 14  
US-09-953-413-18  
Sequence 18, Application US/09953413  
Publication No. US20040018209A1

GENERAL INFORMATION:

APPLICANT: Horwitz, Marcus A.  
TITLE OF INVENTION: Abundant Extracellular  
Products and Methods for Their Production and Use

NUMBER OF SEQUENCES: 91

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kurt A. Maclean  
STREET: 2029 Century Park East, Suite 3800  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.

ZIP: 90067

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0,  
Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/953,413  
FILING DATE: 14-Sep-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION NUMBER: US 08/447,398  
FILING DATE: 23-MAY-1995  
APPLICATION NUMBER: US 08/289,667  
FILING DATE: 12-AUG-1994  
APPLICATION NUMBER: US 08/156,358  
FILING DATE: 23-NOV-1993

ATTORNEY/AGENT INFORMATION:  
NAME: Maclean, Kurt A.  
REGISTRATION NUMBER: 31,118  
REFERENCE/DOCKET NUMBER: 112-272  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (310) 788-5000  
TELEFAX: (310) 277-1297  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 amino acids

TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Mycobacterium tuberculosis  
STRAIN: Erdman  
SEQUENCE DESCRIPTION: SEQ ID NO: 18:  
US-09-953-413-18

Query Match 33.3%; Score 245; DB 3; Length 48;  
Best Local Similarity 100.0%; Pred. No. 2.5e-19;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 35 FDTLRMLRLEDEMKGRYEVRAELPGVDPDQVDIMVRDGLTIKAERT 82  
Db 1 FDTLRMLRLEDEMKGRYEVRAELPGVDPDQVDIMVRDGLTIKAERT 48

RESULT 15  
US-10-147-255-18  
Sequence 18, Application US/10147255  
Publication No. US20030152584A1

GENERAL INFORMATION:

APPLICANT: Horwitz, Marcus A.  
TITLE OF INVENTION: Abundant Extracellular  
Products and Methods for Their Production and Use

NUMBER OF SEQUENCES: 91

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kurt A. Maclean  
STREET: 2029 Century Park East, Suite 3800  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.

ZIP: 90067

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0,  
Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/147,255  
FILING DATE: 15-May-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION NUMBER: US/09/226,539A  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 08/447,398  
FILING DATE: 23-MAY-1995  
APPLICATION NUMBER: US 08/289,667  
FILING DATE: 12-AUG-1994  
APPLICATION NUMBER: US 08/156,358  
FILING DATE: 23-NOV-1993

ATTORNEY/AGENT INFORMATION:  
NAME: Maclean, Kurt A.  
REGISTRATION NUMBER: 31,118  
REFERENCE/DOCKET NUMBER: 112-272  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (310) 788-5000  
TELEFAX: (310) 277-1297  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

```
;
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium tuberculosis
; STRAIN: Erdman
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-10-147-255-18
```

```
Query Match 33.3%; Score 245; DB 4; Length 48;
Best Local Similarity 100.0%; Pred. No. 2.5e-19;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 35 FDTRLMRLEDEMKGRYEVRAELPGVDPDKVDIMVRDGLTIKAERT 82
|||
Db 1 FDTRLMRLEDEMKGRYEVRAELPGVDPDKVDIMVRDGLTIKAERT 48
```

```
Search completed: March 23, 2006, 06:28:12
Job time : 31.2823 secs
```

***This Page Blank (uspto)***

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Biocelelation Ltd.

OM protein - protein search, using SW model

Run on: March 23, 2006, 06:17:23 ; Search time 3.45564 Seconds  
(without alignments)  
1192.746 Million cell updates/sec

Title: US-10-617-038-24

Perfect score: 736

Sequence: 1 MATTLPVQRHRSUPEFSE.....SVAVSEKPTKHIQIRSTN 144

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 169630 seqs, 2862289 residues

Total number of hits satisfying chosen parameters: 169630

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	736	100.0	331	US-11-201-519-4	Sequence 4, Appl
2	736	100.0	368	US-11-201-519-24	Sequence 24, Appl
3	736	100.0	433	US-11-201-519-14	Sequence 14, Appl
4	736	100.0	856	US-11-201-519-12	Sequence 12, Appl
5	144.5	19.6	150	US-11-087-039-2558	Sequence 2558, Ap
6	132.5	18.0	153	US-11-087-039-7830	Sequence 7830, Ap
7	130	17.7	210	US-11-087-039-11233	Sequence 11233, A
8	129	17.5	149	US-11-087-039-11382	Sequence 11382, A
9	129	17.5	155	US-11-036-568A-21407	Sequence 21407, A
10	127	17.3	162	US-11-036-568A-16708	Sequence 16708, A
11	126.5	17.2	154	US-11-087-039-8558	Sequence 8558, Ap
12	126.5	17.2	154	US-11-087-039-10259	Sequence 10259, A
13	126.5	17.2	161	US-11-087-039-10976	Sequence 10976, A
14	125.5	17.1	161	US-11-087-039-11804	Sequence 11804, A
15	124	16.8	159	US-11-036-568A-7013	Sequence 7013, Ap
16	123.5	16.8	153	US-11-087-039-12190	Sequence 12190, A
17	121.5	16.5	131	US-11-036-568A-21408	Sequence 21408, A
18	119.5	16.2	128	US-11-087-039-11932	Sequence 11932, A
19	118.5	16.1	156	US-11-087-039-5432	Sequence 5432, Ap
20	111.5	15.1	150	US-11-087-039-1077	Sequence 1077, Ap
21	110.5	15.0	146	US-11-087-039-11787	Sequence 11787, A
22	109	14.8	141	US-11-074-176-304	Sequence 304, Appl
23	109	14.8	147	US-11-074-176-14	Sequence 14, Appl
24	106.5	14.5	133	US-11-087-039-8784	Sequence 8784, Ap
25	104	14.1	161	US-11-036-568A-14759	Sequence 14759, A

26	104	14.1	198	US-11-036-568A-14758	Sequence 14758, A
27	103	14.0	150	US-11-036-568A-6324	Sequence 6324, Ap
28	103	14.0	153	US-11-036-568A-6323	Sequence 6323, Ap
29	103	14.0	154	US-11-036-568A-6322	Sequence 6322, Ap
30	94.5	12.8	159	US-11-036-568A-7626	Sequence 7626, Ap
31	94.5	12.8	187	US-11-036-568A-7625	Sequence 7625, Ap
32	92	12.5	214	US-11-036-568A-9568	Sequence 9568, Ap
33	92	12.5	239	US-11-036-568A-9567	Sequence 9567, Ap
34	91	12.4	147	US-11-036-568A-21035	Sequence 21035, A
35	88	12.0	124	US-11-036-568A-21036	Sequence 21036, A
36	79	10.7	175	US-10-821-234-1543	Sequence 1543, Ap
37	78.5	10.7	297	US-11-036-568A-18592	Sequence 18592, A
38	78	10.6	136	US-11-036-568A-14425	Sequence 14425, A
39	75.5	10.3	205	US-10-878-556A-199	Sequence 199, Appl
40	74	10.1	210	US-11-010-239-77	Sequence 77, Appl
41	71.5	9.7	242	US-10-467-657-7306	Sequence 7306, Ap
42	71.5	9.7	1401	US-11-036-568A-28512	Sequence 28512, A
43	71	9.6	149	US-11-036-568A-15190	Sequence 15190, A
44	71	9.6	225	US-11-036-568A-15189	Sequence 15189, A
45	70.5	9.6	505	US-11-087-039-2851	Sequence 2851, Ap

## ALIGNMENTS

```

RESULT 1
US-11-201-519-4
: Sequence 4, Application US/11201519
: Publication No. US20060034764A1
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Dillon, Davin C.
: APPLICANT: Alderson, Mark
: APPLICANT: Campos-Neto, Antonio
: TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
: FILE REFERENCE: 014058-009020US
: CURRENT APPLICATION NUMBER: US/11/201,519
: CURRENT FILING DATE: 2005-08-10/359,460
: PRIOR APPLICATION NUMBER: US/10/359,460
: PRIOR FILING DATE: 2003-02-05
: PRIOR APPLICATION NUMBER: US/09/287,849
: PRIOR FILING DATE: 1999-04-07
: PRIOR APPLICATION NUMBER: US 08/818,112
: PRIOR FILING DATE: 1997-03-13
: PRIOR APPLICATION NUMBER: US 08/942,578
: PRIOR FILING DATE: 1997-10-01
: PRIOR APPLICATION NUMBER: US 09/025,197
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 09/056,556
: PRIOR FILING DATE: 1998-04-07
: PRIOR APPLICATION NUMBER: US 09/223,040
: PRIOR FILING DATE: 1998-12-30
: NUMBER OF SEQ ID NOS: 46
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 4
: LENGTH: 331
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence:trt-fusion
US-11-201-519-4

```

```

Query Match 100.0%; Score 736; DB 7; Length 331;
Best Local Similarity 100.0%; Pred. No. 2.2e-68;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 1 MATTLPVQRHRSUPEFSELFAPSPFAGLRPTDRLMRLEDEMKEGRYEVRAELPGV 60
Db 8 MATTLPVQRHRSUPEFSELFAPSPFAGLRPTDRLMRLEDEMKEGRYEVRAELPGV 67

```

Qy 61 DPKDVIDIWRDQGLTKARTKEDKPDGRSEFAVGSFVRTVSLPVGADEDDIKATYDKG 120  
Db 68 DPKDVIDIWRDQGLTKARTKEDKPDGRSEFAVGSFVRTVSLPVGADEDDIKATYDKG 127  
Qy 121 ILTVSAVSESGKPTKHIQIRSTN 144  
Db 128 ILTVSAVSESGKPTKHIQIRSTN 151

## RESULT 2

US-11-201-519-24  
Sequence 24, Application US/11201519  
Publication No. US20060034764A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Alderson, Mark  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens  
FILE REFERENCE: 014058-009020US  
CURRENT APPLICATION NUMBER: US/11/201,519  
CURRENT FILING DATE: 2005-08-10  
PRIOR APPLICATION NUMBER: US/10/359,460  
PRIOR FILING DATE: 2003-02-05  
PRIOR APPLICATION NUMBER: US/09/287,849  
PRIOR FILING DATE: 1999-04-07  
PRIOR APPLICATION NUMBER: US 08/818,112  
PRIOR FILING DATE: 1997-03-13  
PRIOR APPLICATION NUMBER: US 08/942,578  
PRIOR FILING DATE: 1997-10-01  
PRIOR APPLICATION NUMBER: US 09/025,197  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 09/056,556  
PRIOR FILING DATE: 1998-04-07  
PRIOR APPLICATION NUMBER: US 09/223,040  
PRIOR FILING DATE: 1998-12-30  
NUMBER OF SEQ ID NOS: 46  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 24  
LENGTH: 368  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: tetra-fusion  
US-11-201-519-24

Query Match 100.0%; Score 736; DB 7; Length 368;

Best Local Similarity 100.0%; Pred. No. 2,5e-68;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATTLPVQRHPRSLFPFSESLFAAPSPAGLRPTFTDRLMKRLDEMKEGRYEVAELPGV 60  
Db 9 MATTLPVQRHPRSLFPFSESLFAAPSPAGLRPTFTDRLMKRLDEMKEGRYEVAELPGV 68  
Qy 61 DPKDVIDIWRDQGLTKARTKEDKPDGRSEFAVGSFVRTVSLPVGADEDDIKATYDKG 120  
Db 69 DPKDVIDIWRDQGLTKARTKEDKPDGRSEFAVGSFVRTVSLPVGADEDDIKATYDKG 128  
Qy 121 ILTVSAVSESGKPTKHIQIRSTN 144  
Db 129 ILTVSAVSESGKPTKHIQIRSTN 152

## RESULT 3

US-11-201-519-14  
Sequence 14, Application US/11201519  
Publication No. US20060034764A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.  
APPLICANT: Alderson, Mark  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens  
FILE REFERENCE: 014058-009020US  
CURRENT APPLICATION NUMBER: US/11/201,519  
CURRENT FILING DATE: 2005-08-10  
PRIOR APPLICATION NUMBER: US/10/359,460  
PRIOR FILING DATE: 2003-02-05  
PRIOR APPLICATION NUMBER: US/09/287,849  
PRIOR FILING DATE: 1999-04-07  
PRIOR APPLICATION NUMBER: US 08/818,112  
PRIOR FILING DATE: 1997-03-13  
PRIOR APPLICATION NUMBER: US 08/942,578  
PRIOR FILING DATE: 1997-10-01  
PRIOR APPLICATION NUMBER: US 09/025,197  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 09/056,556  
PRIOR FILING DATE: 1998-04-07  
PRIOR APPLICATION NUMBER: US 09/223,040  
PRIOR FILING DATE: 1998-12-30  
NUMBER OF SEQ ID NOS: 46  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 14  
LENGTH: 433  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: tetra-fusion  
US-11-201-519-14

Query Match 100.0%; Score 736; DB 7; Length 433;

Best Local Similarity 100.0%; Pred. No. 3.1e-68;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATTLPVQRHPRSLFPFSESLFAAPSPAGLRPTFTDRLMKRLDEMKEGRYEVAELPGV 60  
Db 9 MATTLPVQRHPRSLFPFSESLFAAPSPAGLRPTFTDRLMKRLDEMKEGRYEVAELPGV 68  
Qy 61 DPKDVIDIWRDQGLTKARTKEDKPDGRSEFAVGSFVRTVSLPVGADEDDIKATYDKG 120  
Db 69 DPKDVIDIWRDQGLTKARTKEDKPDGRSEFAVGSFVRTVSLPVGADEDDIKATYDKG 128  
Qy 121 ILTVSAVSESGKPTKHIQIRSTN 144  
Db 129 ILTVSAVSESGKPTKHIQIRSTN 152

## RESULT 4

US-11-201-519-12  
Sequence 12, Application US/11201519  
Publication No. US20060034764A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Alderson, Mark  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens  
FILE REFERENCE: 014058-009020US  
CURRENT APPLICATION NUMBER: US/11/201,519  
CURRENT FILING DATE: 2005-08-10  
PRIOR APPLICATION NUMBER: US/10/359,460  
PRIOR FILING DATE: 2003-02-05  
PRIOR APPLICATION NUMBER: US/09/287,849  
PRIOR FILING DATE: 1999-04-07  
PRIOR APPLICATION NUMBER: US 08/818,112  
PRIOR FILING DATE: 1997-03-13  
PRIOR APPLICATION NUMBER: US 08/942,578

```
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 12
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: penta-fusion
US-11-201-519-12
```

```
Query Match      100.0%; Score 736; DB 7; Length 856;
Best Local Similarity 100.0%; Pred. No. 7,8e-68;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 MATTLPVORHPRSLPPEFSELPAAFPSPAGLRPTFTRLMRLEDEMKGRYEVRAELPGV 60
DB      9 MATTLPVORHPRSLPPEFSELPAAFPSPAGLRPTFTRLMRLEDEMKGRYEVRAELPGV 68
QY      61 DDDKVDIMVRDQGLTIKAERTQKDFGRSEFAYGSFVRTVSLPVGADEDDIKATYDKG 120
DB      69 DDDKVDIMVRDQGLTIKAERTQKDFGRSEFAYGSFVRTVSLPVGADEDDIKATYDKG 128
QY      121 ILTVSVAVSEKGPTEKHIOIRSTN 144
DB      129 ILTVSVAVSEKGPTEKHIOIRSTN 152
```

```
RESULT 5
US-11-087-099-2558
; Sequence 2558, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 2558
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Hordeum vulgare subsp. vulgare
US-11-087-099-2558
```

```
Query Match      19.6%; Score 144.5; DB 7; Length 150;
Best Local Similarity 32.1%; Pred. No. 7,6e-08;
Matches 45; Conservative 27; Mismatches 53; Indels 15; Gaps 5;
```

```
QY      15 FPEFSELPFA-----AFSPFAGLRPTFTRLMRLEDEMKGRYEVRAELPGVDPDDKV 66
DB      10 FDPFADLWADPDTFRSIVPAFSGNSFTAFANAAVDWKETPEAHVFAFDLPGVKKE-EV 68
QY      67 DIMVRDQGLTIKAERTQKDFDGRSEFAYGSFVRTVSLPVGADEDDIKATYDKGI 121
DB      69 KVEVEDGANVLVSGRTEKEDKNDKMRVERSSGKFTYRFRLLPEDAKVEEVKAGLENGV 128
QY      122 LTVSVAVSE-GKPTKHIQI 140
DB      129 LTVTVKTEVKKPEVKAIEI 148
```

```
RESULT 6
US-11-087-099-7830
; Sequence 7830, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 7830
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Funaria hygrometrica
US-11-087-099-7830
```

```
Query Match      18.0%; Score 132.5; DB 7; Length 153;
Best Local Similarity 38.7%; Pred. No. 1,4e-06;
Matches 36; Conservative 18; Mismatches 32; Indels 7; Gaps 4;
```

```
QY      54 RABLPGVDPDQVIMVRDQGLTIKAERTQKDFGRSEFAYGSFVRTVSLPVGA 108
DB      61 KADLPGLTKE-EVKVQLGRTLEICGRKKEVQKSDTWMRBAQGSFMRRLPBG 119
QY      109 DEDDIKATYDKGILTVSV-AVSEKGPTEKHIOI 140
DB      120 NTDDVKAGVQDQGLTVTVPKVQKPKQVQRIEI 152
```

```
RESULT 7
US-11-087-099-11233
; Sequence 11233, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 11233
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Papaver somniferum
US-11-087-099-11233
```

```
Query Match      17.7%; Score 130; DB 7; Length 210;
Best Local Similarity 30.5%; Pred. No. 3,8e-06;
Matches 46; Conservative 23; Mismatches 58; Indels 24; Gaps 6;
```

```
QY      13 SLPEFES-----ELPAFPSPAG-----LRPTFTRLMRLEDEMKGRYEVRAE 56
DB      14 NVFPFSLDIWDPOGFEPFSTGALTANWQGSSTARETSQLANTRIDWKETPEAHVFPAD 73
QY      57 LPGVDPDQVIMVRDQGLTIKAERTQKDFGRSEFAYGSFVRTVSLPVGADE 110
DB      74 LPTVTKKE-EVAVVEBGRVLDIISGRSRSEBKNDKMRVERSSGKTLRFRLLPENTKM 132
QY      111 DDIKATYDKGILTVSV-AVSEKGPTEKHIOI 140
DB      133 DEVKATMENGVLTVCVKEVQRRREVKSIEI 163
```

```
RESULT 8
US-11-087-099-11382
; Sequence 11382, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 11382
; LENGTH: 149
; TYPE: PRT
```

```

; ORGANISM: Oryza sativa (Japonica cultivar-group)
US-11-087-099-11382
Query Match      17.5%; Score 129; DB 7; Length 149;
Best Local Similarity 33.3%; Pred. No. 3e-06;
Matches 45; Conservative 24; Mismatches 48; Indels 18; Gaps 7;

Qy      12 RSLPFSESLFAPPSFAGLRPTFDTRLRLDEMEKGRYEVRAELPGVDKDVIMVR 71
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      25 RSLVPATSDRDTA--AFANARVDW-----KETPES-HVFKADLPVKKE-EVKVEVE 72

Qy      72 DGO-LTIKERTEQKDFG-----RSEFAYGFVRTVSLPVGADDDIKATYDKGILITVSV 126
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      73 EGNVLVISGGRSKKEDKNDKMHVERSSGQFMKRFRLPENAKVDQVAKSMENGVLITV 132

Qy      127 AVSE-GKPTKHIQI 140
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      133 PKAEVKKPEVKAIEI 147

RESULT 9
US-11-096-568A-21407
; Sequence 21407, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 21407
; LENGTH: 155
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; NAME/KEY: misc_feature
; LOCATION: (1)..(155)
; OTHER INFORMATION: Cerees Seq. ID no. 12403861
US-11-096-568A-21407

Query Match      17.5%; Score 129; DB 7; Length 155;
Best Local Similarity 29.8%; Pred. No. 3.2e-06;
Matches 45; Conservative 27; Mismatches 61; Indels 18; Gaps 6;

Qy      7 VQHRPSLFPFES-ELFAPPSFAGLRP-----TFDTRLRLDEMEKGRYEVRA 55
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      4 VSRKSSNVDFPSLDLMDPDMFRSIVPSAASSGGSETAFANARVDKETPREAHYFKA 63

Qy      56 ELPGVDKDVIMVRDGOQLT---KARTEQKDFD--GRSEFAYGSFVRTVSLPVGAD 110
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      64 DLPEIKKE-EVKVEVEDGNVLVISGKRSREEDKNDKMHVERSSGQFMKRFRLPENAKV 122

Qy      111 DDITATYDKGILITVSAVSE-GKPTKHIQI 140
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      123 DEVKAGLENGVLITVTPKTEYKKEPVKAIEI 153

RESULT 10
US-11-096-568A-16708
; Sequence 16708, Application US/11096568A
; Publication No. US2006004240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 16708
; LENGTH: 162
; TYPE: PRT

```

```

; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(162)
; OTHER INFORMATION: Cerees Seq. ID no. 12353810
US-11-096-568A-16708

Query Match      17.3%; Score 127; DB 7; Length 162;
Best Local Similarity 31.2%; Pred. No. 5.4e-06;
Matches 39; Conservative 25; Mismatches 51; Indels 10; Gaps 5;

Qy      25 PPSFAGLRPTFDTRLM---RLDEMEKGRYEVRAELPGVDKDVIMVRDGO-LTIKAE 80
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      37 PPSFGGTTTSSETAFAFAGARVDWKETPEAHVFTKDVGKKKE-EVKVELEDGNVLQISGE 95

Qy      81 RT-----EQKDFDGRSEFAYGSFVRTVSLPVGADDDIKATYDKGILITVSAVSEGPTE- 135
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      96 RSKQEERKDTWHRVERSSGKFLRRFLPENARTEQLSAEMENGVLITVTPKEAKKADV 155

Qy      136 KHIQI 140
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      156 KSIQI 160

RESULT 11
US-11-087-099-8558
; Sequence 8558, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 8558
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Quercus suber
US-11-087-099-8558

Query Match      17.2%; Score 126.5; DB 7; Length 154;
Best Local Similarity 31.9%; Pred. No. 5.7e-06;
Matches 46; Conservative 24; Mismatches 55; Indels 19; Gaps 7;

Qy      13 SLPFES-----ELFAPPSFAGLRP-----TFDTRLRLDEMEKGRYEVRAELPGVDP 62
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      12 NVDPPESLDIDMPREGSAVASVPSARETTATAT--ARIDWKETPREAHIFKADLPOLKK 69

Qy      63 KDVIMVRDGO-LTIKERT---EQKDFDGRSEFAYGSFVRTVSLPVGADDDIKATY 117
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      70 E-EVKVEVEDGNVLQISGERSKEHEEKNDKMHVERSCGFMKRFRLPENAKVDQVAKNM 128

Qy      118 DKGILITVSAVSE-GKPTKHIQI 140
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      129 ENGVLITVMPKEBOKKPAVAIAEI 152

RESULT 12
US-11-087-099-10259
; Sequence 10259, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 10259
; LENGTH: 154
; TYPE: PRT

```





***This Page Blank (uspto)***

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 04:56:29 ; Search time 88.0019 Seconds  
(without alignments)  
1652.628 Million cell updates/sec

Title: US-10-617-038-25  
Perfect score: 1718  
Sequence: 1 MPDTWTTDVIKSAVOLACR.....PATPRRPIDEVFHAKDHR 331

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues  
Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*  
9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1718	100.0	331	8 ADI37304	ADI37304 M. tuberc
2	915	53.3	344	8 ADI37315	ADI37315 M. tuberc
3	529	30.8	104	2 AAY39048	AAY39048 M. tuberc
4	529	30.8	104	2 AAY39191	AAY39191 M. tuberc
5	417	24.3	332	8 ADI37319	ADI37319 M. tuberc
6	106.5	6.2	413	3 AAB32537	AAB32537 S. lavend
7	106.5	6.2	413	7 ADE10309	ADE10309 S. lavend
8	104	6.1	380	6 ABU20272	ABU20272 Protein e
9	100.5	5.8	1176	4 AAB70151	AAB70151 DNA encod
10	100.5	5.8	1176	4 AAM79269	AAM79269 Human pro
11	100.5	5.8	1176	8 ADL13064	ADL13064 Human ste
12	99	5.8	897	8 ADF54681	ADF54681 Aconitase
13	97.5	5.7	459	4 AAU48145	AAU48145 Propionib
14	97.5	5.7	459	4 ABM44664	ABM44664 Propionib
15	97	5.6	585	5 ABP69805	ABP69805 Human pol
16	97	5.6	904	8 ADF54679	ADF54679 Aconitase
17	96.5	5.6	423	6 ABP99222	ABP99222 Orchomy
18	95.5	5.6	3956	8 ADV99898	ADV99898 Nanchangm
19	95	5.5	534	7 ADB65249	ADB65249 Human pro
20	95	5.5	543	6 ADA54862	ADA54862 Human pro
21	95	5.5	543	9 ADX06434	ADX06434 Cyc1in-de
22	95	5.5	585	8 ADRO9359	ADRO9359 Human pro
23	95	5.5	585	8 ADU82609	ADU82609 Human MDD
24	94	5.5	378	6 ABU39660	ABU39660 Protein e

25	94	5.5	437	7 ABM66609	ABM66609 Rice abio
26	93.5	5.4	188	3 AAY81711	AAY81711 Streptoco
27	93.5	5.4	201	6 ABU00998	ABU00998 S. pneumo
28	93.5	5.4	201	8 ADM92111	ADM92111 S. pneumo
29	93	5.4	431	7 ABM85743	ABM85743 Human pro
30	93	5.4	932	6 ABU34318	ABU34318 Protein e
31	92.5	5.4	519	7 ABO82069	ABO82069 Pseudomon
32	92.5	5.4	737	8 ADJ93777	ADJ93777 AChE1 pro
33	92.5	5.4	7349	6 ABU11385	ABU11385 Protein e
34	92	5.4	200	5 ABP26703	ABP26703 Streptoco
35	92	5.4	200	8 ADV88610	ADV88610 Streptoco
36	92	5.4	200	8 ADV79863	ADV79863 Streptoco
37	92	5.4	200	8 ADV82013	ADV82013 Streptoco
38	92	5.4	333	5 ABP66056	ABP66056 BtFidobac
39	91.5	5.3	750	6 ABU36645	ABU36645 Protein e
40	91.5	5.3	750	6 ABU34712	ABU34712 Protein e
41	91.5	5.3	1346	2 AAM31346	AAM31346 Rat tumou
42	90.5	5.3	457	6 ABU35824	ABU35824 Protein e
43	90.5	5.3	664	7 ABO68223	ABO68223 Pseudomon
44	90.5	5.3	737	8 ADJ93781	ADJ93781 AChE1 pro
45	90.5	5.3	841	6 ABU26085	ABU26085 Protein e

ALIGNMENTS

RESULT 1  
ADI37304 standard; protein; 331 AA.  
ADI37304;  
ADI37304;  
22-APR-2004 (first entry)  
M. tuberculosis low oxygen induced antigen Rv2032 SEQ ID NO:25.  
mycobacterial infection; vaccine; tuberculosis;  
Mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;  
low oxygen induced antigen.  
Mycobacterium tuberculosis.  
WO2004006952-A2.  
22-JAN-2004.  
08-JUL-2003; 2003WO-DK000477.  
13-JUL-2002; 2002DK-00001098.  
(STAT-) STATENS SERUM INST.  
Andersen P, Rosenkrands I, Stryhn A,  
WPI; 2004-122778/12.  
N-PDB; ADI37349.  
Use of one or more polypeptides or their fragments, which are expressed during the latent stage of the mycobacterial infection, and/or nucleic acids encoding the polypeptides, for a therapeutic vaccine against tuberculosis.  
Claim 3, SEQ ID NO 25; 76pp; English.  
The present invention describes polypeptides or their fragments, which are expressed during the latent stage of a mycobacterial infection, and/or nucleic acids encoding the polypeptides, which are useful for creating a therapeutic vaccine against tuberculosis. Also described: (1) a therapeutic vaccine against tuberculosis comprising one or more polypeptides; (2) a method for creating an animal, including a human being, with tuberculosis caused by virulent mycobacteria, e.g. by Mycobacterium tuberculosis, M. africanum or M. bovis; (3) a method for immunising an animal, including a human being, against tuberculosis

CC caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis  
CC caused by virulent mycobacteria in an animal, including a human being;  
CC (5) a method for diagnosing previous or ongoing infection with a virulent  
CC mycobacterium; and (6) a method of diagnosing Mycobacterium tuberculosis  
CC infection in a subject. The polypeptides have antibacterial activities,  
CC and can be used in vaccines and in gene therapy. The polypeptides are  
CC useful for the manufacture of a therapeutic vaccine for treating an  
CC individual who is infected by a virulent mycobacterium, e.g. M.  
CC tuberculosis, and who is not vaccinated with BCG against tuberculosis.  
CC The present sequence represents a low oxygen induced antigen, which is  
CC used in the exemplification of the present invention.

XX Sequence 331 AA;

Query Match 100.0%; Score 1718; DB 8; Length 331;  
Best Local Similarity 100.0%; Pred. No. 4,1e-172; Mismatches 0; Gaps 0;  
Matches 331; Conservative 0; Indels 0; Gaps 0;

QY 1 MPDTMTTVDVKSAYQLACRAPSLHNSQPMRWIAEDHTVALFLDKRVLVATDHSGRAL 60  
DB 1 MPDTMTTVDVKSAYQLACRAPSLHNSQPMRWIAEDHTVALFLDKRVLVATDHSGRAL 60  
QY 61 LGGCAVLDHFRVAAAAGTTANVERFPNPNDPLASIDESPADPTGHRLRADAILLR 120  
DB 61 LGGCAVLDHFRVAAAAGTTANVERFPNPNDPLASIDESPADPTGHRLRADAILLR 120  
QY 121 RTDLRPAEPDMDLVESQRLTATTADTVRIDVADMRPELAASKLTESLRLYDSYH 180  
DB 121 RTDLRPAEPDMDLVESQRLTATTADTVRIDVADMRPELAASKLTESLRLYDSYH 180  
QY 121 RTDLRPAEPDMDLVESQRLTATTADTVRIDVADMRPELAASKLTESLRLYDSYH 180  
DB 121 RTDLRPAEPDMDLVESQRLTATTADTVRIDVADMRPELAASKLTESLRLYDSYH 180  
QY 181 AELFWMTGAFETSGIPHSLSVSAESDRVTGGRDPVANTDRPFGHRSKVLVLT 240  
DB 181 AELFWMTGAFETSGIPHSLSVSAESDRVTGGRDPVANTDRPFGHRSKVLVLT 240  
QY 241 YDNERASLRCGEMLSAVLDTMAAGLATCTLTHITELHASRDVLAALIGQAPQALVR 300  
DB 241 YDNERASLRCGEMLSAVLDTMAAGLATCTLTHITELHASRDVLAALIGQAPQALVR 300  
QY 301 VGLAPEMEPPPATPRRPIDEVFHVAKDHR 331  
DB 301 VGLAPEMEPPPATPRRPIDEVFHVAKDHR 331

RESULT 2  
AD137315  
ID AD137315 standard; protein: 344 AA.

XX AD137315;

XX 22-APR-2004 (first entry)

DE M. tuberculosis low oxygen induced antigen Rv3127 SEQ ID NO:36.

KM mycobacterial infection; vaccine; tuberculosis;

KM mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;

XX low oxygen induced antigen.

OS Mycobacterium tuberculosis.

XX WO2004006952-A2.

XX 22-JAN-2004.

XX 08-JUL-2003; 2003WO-DK00477.

XX 13-JUL-2002; 2002DK-00001098.

XX (STAT-) STATENS SERUM INST.

XX Andersen P, Rosenkrands I, Stryhn A;

XX WPI, 2004-122778/12.

XX N-PSDB; AD137360.

XX Use of one or more polypeptides or their fragments, which are expressed  
PT during the latent stage of the mycobacterial infection, and/or nucleic  
PT acids encoding the polypeptides, for a therapeutic vaccine against  
PT tuberculosis.

XX Claim 3; SEQ ID NO 36; 76pp; English.

XX The present invention describes polypeptides or their fragments, which  
CC are expressed during the latent stage of a mycobacterial infection,  
CC and/or nucleic acids encoding the polypeptides, which are useful for  
CC creating a therapeutic vaccine against tuberculosis. Also described: (1)  
CC a therapeutic vaccine against tuberculosis comprising one or more  
CC polypeptides; (2) a method for treating an animal, including a human  
CC being, with tuberculosis caused by virulent mycobacteria, e.g. by  
CC immunising an animal, including a human being, against tuberculosis  
CC caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis  
CC caused by virulent mycobacteria in an animal, including a human being;  
CC (5) a method for diagnosing previous or ongoing infection with a virulent  
CC mycobacterium; and (6) a method of diagnosing Mycobacterium tuberculosis  
CC infection in a subject. The polypeptides have antibacterial activities,  
CC and can be used in vaccines and in gene therapy. The polypeptides are  
CC useful for the manufacture of a therapeutic vaccine for treating an  
CC individual who is infected by a virulent mycobacterium, e.g. M.  
CC tuberculosis, and who is not vaccinated with BCG against tuberculosis.  
CC The present sequence represents a low oxygen induced antigen, which is  
CC used in the exemplification of the present invention.

XX Sequence 344 AA;

Query Match 53.3%; Score 915; DB 8; Length 344;  
Best Local Similarity 56.7%; Pred. No. 2.6e-87; Mismatches 182; Conservative 39; Indels 96; Gaps 2;

QY 10 VISAQVLACRAPSLHNSQPMRWIAE---DH-VVALFLDKRVLVATDHSGRALGCGA 65  
DB 1 VLNKAVALACRAPSVHNSQPMRWVASGSEHTTVHLFVNHRVTPADHSGRAALISCGA 60  
QY 66 VLDHFRVAAAAGTTANVERFPNPNDPLASIDESPADPTGHRLRADAILLRTRDL 125  
DB 61 VLDHFRVAAAAGTTANVERFPNPNDPLASIDESPADPTGHRLRADAILLRTRDL 120  
QY 126 PFAEPDMDLVESQRLTATTADTVRIDVADMRPELAASKLTESLRLYDSYHAELEW 185  
DB 121 PFDSPMYHLEFPALRDVADVKAEMLDVSDQRTLRVVASQSEVLRRDDPYHALEW 180  
QY 186 WTGAFETSGIPHSLSVSAESDRVTGGRDPVANTDRPFGHRSKVLVLTSDNER 245  
DB 181 WTSPEVLAHGVPPDTLASDAERLVRVLDGRFPVRSYQNRRAELADRSKVLVLTSDNR 240  
QY 246 ASLRCGEMLSAVLDTMAAGLATCTLTHITELHASRDVLAALIGQAPQALVRGVAL 305  
DB 241 ADALRCGEVSTILLECTMAAGLATCTLTHITELHASRDVLAALIGQAPQALVRGVAL 300  
QY 306 EMEPPPATPRRPIDEVFHV 326  
DB 301 FLAAVPAPTRPRRPLDSVLQIR 321

RESULT 3  
AAV39048  
ID AAV39048 standard; protein: 104 AA.

XX AAV39048;

XX 05-NOV-1999 (first entry)

DE M. tuberculosis recombinant antigen protein LSR-12.

KM Antigen; diagnosis; detection; infection; antibody; immunisation;

XX vaccine; immunity.

```

OS   Mycobacterium tuberculosis.
XX
XX   WO9942118-A2.
XX
XX   26-AUG-1999.
XX
XX   17-FEB-1999; 99WO-US003265.
XX
XX   18-FEB-1998; 98US-00024753.
XX   05-MAY-1998; 98US-00072596.
XX
XX   (CORI-) CORIXA CORP.
XX
XX   Reed SG, Skeiky YAM, Dillon DC, Campos-Neto A, Houghton R;
XX   Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX   WPI; 1999-527416/44.
XX   N-PSDB; AAZ19173.
XX
XX   New polypeptide comprising antigenic portions of M. tuberculosis.
XX
XX   Example 3D; Page 268-269; 323pp; English.
XX
XX   This invention describes novel recombinant antigens and their encoding
XX   nucleic acids derived from Mycobacterium tuberculosis. The novel
XX   polypeptides are useful for detecting M. tuberculosis infection in a
XX   biological sample by detecting antibodies which bind with the
XX   polypeptides, and are useful as vaccines for immunizing against M.
XX   tuberculosis infection. The new detection methods are needed as current
XX   vaccination strategies do not provide 100% immunity
XX
XX   Sequence 104 AA;
XX
XX   Query Match      30.8%; Score 529; DB 2; Length 104;
XX   Best Local Similarity 100.0%; Pred. No. 2.5e-47;
XX   Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX   74   MAAAGTTANVERFPNPNDPLHLASIDFSPADFTVTEGHRLRADAILLRTRDRLPFAEPPDW 133
XX   1   MAAAGTTANVERFPNPNDPLHLASIDFSPADFTVTEGHRLRADAILLRTRDRLPFAEPPDW 60
XX
XX   134   DLVESQRLRTVTYADTVRIDVIADDMRPDLAASKLTESLRLYDS 177
XX   61   DLVESQRLRTVTYADTVRIDVIADDMRPDLAASKLTESLRLYDS 104
XX
XX   Db
XX
XX   RESULT 4
XX   AAY39191
XX   ID   AAY39191 standard; protein; 104 AA.
XX
XX   AC   AAY39191;
XX
XX   DT   05-NOV-1999 (first entry)
XX
XX   DE   M. tuberculosis antigen L5ER-12 amino acid sequence.
XX
XX   KW   Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
XX   KW   immunotherapy; diagnosis; immunisation; vaccine; infection;
XX   KW   immune response; skin test.
XX
XX   OS   Mycobacterium tuberculosis.
XX
XX   PN   WO9942076-A2.
XX
XX   PD   26-AUG-1999.
XX
XX   PF   17-FEB-1999; 99WO-US003268.
XX
XX   PR   18-FEB-1998; 98US-00025197.
XX   05-MAY-1998; 98US-00072967.
XX
XX   (CORI-) CORIXA CORP.
XX

```

```

PI   Reed SG, Skeiky YAM, Dillon DC, Campos-Neto A, Houghton R;
XX   Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX   WPI; 1999-527409/44.
XX   N-PSDB; AAZ19385.
XX
XX   New antigens from Mycobacterium tuberculosis useful in diagnostic skin
XX   tests and protective or therapeutic vaccines or compositions.
XX
XX   Example 3; Page 223; 299pp; English.
XX
XX   The present invention describes polypeptides comprising an immunogenic
XX   part of a Mycobacterium tuberculosis antigen (Ag). Also described are
XX   vaccines and fusion protein containing M. tuberculosis Ag's. M.
XX   tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
XX   polypeptides fragments, can be used in pharmaceutical compositions or
XX   vaccines to generate a protective or therapeutic immune response to M.
XX   tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
XX   Ag can induce proliferation of, or cytokine secretion by, T, B or natural
XX   killer cells and/or macrophages in tuberculosis-immune subjects. AAZ19249
XX   to AAZ19460 and AAY39083 to AAY39225 are used in the exemplification of
XX   the present invention
XX
XX   Sequence 104 AA;
XX
XX   Query Match      30.8%; Score 529; DB 2; Length 104;
XX   Best Local Similarity 100.0%; Pred. No. 2.5e-47;
XX   Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX   74   MAAAGTTANVERFPNPNDPLHLASIDFSPADFTVTEGHRLRADAILLRTRDRLPFAEPPDW 133
XX   1   MAAAGTTANVERFPNPNDPLHLASIDFSPADFTVTEGHRLRADAILLRTRDRLPFAEPPDW 60
XX
XX   134   DLVESQRLRTVTYADTVRIDVIADDMRPDLAASKLTESLRLYDS 177
XX   61   DLVESQRLRTVTYADTVRIDVIADDMRPDLAASKLTESLRLYDS 104
XX
XX   Db
XX
XX   RESULT 5
XX   ADI37319
XX   ID   ADI37319 standard; protein; 332 AA.
XX
XX   AC   ADI37319;
XX
XX   DT   22-APR-2004 (first entry)
XX
XX   DE   M. tuberculosis low oxygen induced antigen Rv3131 SEQ ID NO:40.
XX
XX   KW   mycobacterial infection; vaccine; tuberculosis;
XX   KW   Mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;
XX   KW   low oxygen induced antigen.
XX
XX   OS   Mycobacterium tuberculosis.
XX
XX   PN   WO2004006952-A2.
XX
XX   PD   22-JAN-2004.
XX
XX   PF   08-JUL-2003; 2003WO-DK000477.
XX
XX   PR   13-JUL-2002; 2002DK-00001098.
XX
XX   (START-) STATENS SERUM INST.
XX
XX   PI   Andersen P, Rosenkrands I, Stryhn A;
XX   WPI; 2004-122778/12.
XX   N-PSDB; ADI37364.
XX
XX   Use of one or more polypeptides or their fragments, which are expressed
XX   during the latent stage of the mycobacterial infection, and/or nucleic
XX   acids encoding the polypeptides, for a therapeutic vaccine against
XX   tuberculosis.
XX

```

XX Claim 3; SEQ ID NO 40; 76pp; English.  
XX  
XX The present invention describes polypeptides or their fragments, which  
CC are expressed during the latent stage of a mycobacterial infection,  
CC and/or nucleic acids encoding the polypeptides, which are useful for  
CC creating a therapeutic vaccine against tuberculosis. Also described: (1)  
CC a therapeutic vaccine against tuberculosis comprising one or more  
CC polypeptides; (2) a method for treating an animal, including a human  
CC being, with tuberculosis caused by virulent mycobacteria, e.g. by  
CC Mycobacterium tuberculosis, M. africanum or M. bovis; (3) a method for  
CC immunising an animal, including a human being, against tuberculosis  
CC caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis  
CC caused by virulent mycobacteria in an animal, including a human being;  
CC (5) a method for diagnosing previous or ongoing infection with a virulent  
CC mycobacterium; and (6) a method of diagnosing Mycobacterium tuberculosis  
CC infection in a subject. The polypeptides have antibacterial activities,  
CC and can be used in vaccines and in gene therapy. The polypeptides are  
CC useful for the manufacture of a therapeutic vaccine for treating an  
CC individual who is infected by a virulent mycobacterium, e.g. M.  
CC tuberculosis, and who is not vaccinated with BCG against tuberculosis.  
CC The present sequence represents a low oxygen induced antigen, which is  
CC used in the exemplification of the present invention.

XX Sequence 332 AA;

Query Match 24.3%; Score 417; DB 8; Length 332;  
Best Local Similarity 36.1%; Pred. No. 9.5e-35;  
Matches 120; Conservative 45; Mismatches 139; Indels 28; Gaps 11;

QY PDVTWTTDVYKSAVOLACRAPSLNSQPMWIAEDHTVALFLDQRYVATDHSGRRL 61  
D 6 PD---AETVATVTLAVRAPSHNTQPMRWKVCPTSLSPDPMQSRSTDPGRELIL 61  
QY GCGAVLDHFRVMAAAGTTAVERPNDPLHLASIDFSP-----ADFTVGHRLRADA 116  
D 62 SCGVALHCHVVALSLGQAVKVRFPDCKDCHLTIIVQVLVDPQADVAL-----AAA 115  
QY 117 ILRLRTDRLPPA--EPPDMDLVESQRLTQTVADTVRIDVADMRPELAAASKLTESRL 174  
D 116 IPRRTDRBAYSCWVPFGDIALMAARAARGWMLR-QVSLDNRKAIQAVALD---HV 171  
QY 175 YDSYHAELFWMTGAFETSEGIPIHSSLVSAESDRVTGGRF--PVAN-TDRRPERGHD 231  
D 172 TDEYLRBELTWSRGYGVAGVPAKNEDPSAPIP-GRLFAGGGLSQPSDVLV-P-ADD 228  
QY 232 RSKVLVSTYDNESASLRCGEMLSAVYLDATMGATCTTHTHELASDLV-AALIG 290  
D 229 GAAILAGTETDDRLARLRAGEAASIVLITAMGGLACCPITEPELIXTRDAVRAVEFG 288  
QY 291 QPATQALVAVGLAPEMEPEPPATPRRPIDEV 322  
D 289 AGGYFQMLLRVGMAPINADPLPPTPRRELQOV 320

RESULT 6  
AAB32537  
ID AAB32537 standard; protein; 413 AA.

XX AAB32537;  
XX  
XX 19-JAN-2001 (first entry)  
XX  
XX S. lavendulae Mmex encoded protein sequence.  
XX  
XX Mitomycin, biosynthesis; mitosome ring system; antibiotic; anti-cancer;  
XX anti-inflammatory; immune-enhancer; immunosuppressant; asthma;  
XX chronic obstructive pulmonary disease; respiratory inflammation;  
XX fungicide; pesticide.  
XX  
XX Streptomyces lavendulae.  
XX  
XX WO200053737-A2.

XX 14-SEP-2000.  
XX  
XX 10-MAR-2000; 2000WO-US006394.  
XX  
XX 12-MAR-1999; 99US-0026965.  
XX  
XX (MNU) UNIV MINNESOTA.  
XX (SHER/) SHERMAN D H.  
XX (MAOY/) MAO Y.  
XX (VARO/) VAROGLU M.  
XX (HEM/) HE M.  
XX (SHEL/) SHELDON P C.  
XX  
XX Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;  
XX WPI; 2000-601980/57.  
XX N-PSDB; AAC55838.  
XX  
XX Novel nucleic acid molecule comprising mitomycin biosynthetic gene  
XX cluster useful for cloning mitomycin biosynthetic genes for elucidating  
XX the molecular basis of mitosome ring system biosynthesis.  
XX  
XX Disclosure; Page 390-392; 399pp; English.

XX This invention relates to isolated and purified nucleic acid molecules  
XX from the mitomycin biosynthetic gene cluster. Mitomycins are a group of  
XX natural products that contain a variety of functional groups, including  
XX amino benzquinone and axiridine ring systems. The S. lavendulae  
XX mitomycin biosynthetic gene cluster comprises 47 mitomycin genes spanning  
XX 55kb of DNA. The invention includes an expression cassette comprising a  
XX mitomycin biosynthetic gene operably linked to a promoter, and host cells  
XX transformed with the cassette. The nucleotide, and protein sequences and  
XX the transformed host cells of the invention result in antitastmatic,  
XX antitumorigenic, cytotoxic, immunomodulatory, and antibiotic  
XX activities. The nucleotide sequences are used to elucidate the molecular  
XX basis for the biosynthesis of the mitosome ring system, as well as to  
XX engineer the biosynthesis of novel natural products, e.g. antibiotics,  
XX anti-inflammatory agents, anti-cancer agents, immune-enhancers,  
XX immunosuppressants, agents to treat asthma, chronic obstructive pulmonary  
XX disease as well as other disease involving respiratory inflammation, or  
XX cholesterol-lowering agents or as crop protection agents (e.g. fungicides  
XX or insecticides) as well as biopolymers, e.g., in packaging or biomedical  
XX applications, or to engineer PHA monomer synthases. Sequences AAC5782-  
XX C55881, AAC55815-C55849 and AAB32485-B32542 represent mitomycin  
XX biosynthetic gene cluster DNA sequences and encoded proteins. Sequences  
XX AAC55812-C55814, AAC55850-C55856 and AAC55862-C55869 represent PCR  
XX primers used in the cloning of the mitomycin biosynthetic genes

XX Sequence 413 AA;

Query Match 6.2%; Score 106.5; DB 3; Length 413;  
Best Local Similarity 24.1%; Pred. No. 0.081;  
Matches 73; Conservative 32; Mismatches 111; Indels 87; Gaps 14;

QY 2 PDVTWTTDVYKSAVOLACRAPSLNSQPMWIAEDHTVALFLDKDR-----VIXAT 52  
D 107 PDPHYADPRLRGVIRQRYPDIRHPLVW-----NELKGFYDEDRRRMDEGYETRLYNL 160  
QY 53 DH-----SGREALGCGAVLDHFRVMAAAG-----TTAVERPNDNDP 92  
D 161 VHAELKRRNPNLVGGGYAVVDDHPPEDDAADSRRLRGWGBLQDQSAVIRYKN----- 216  
QY 93 LHLASIDFSPAD---FVTEGRLRADAILRLTRDRLFAEPDMDLVESQLRTTVTADTV 149  
D 217 AHKAGADVVDSSYTRBCHRAIPDS--FAATK--FAVTVM--VRSTGLPVWMAEW 270  
QY 150 RIIDVIADKRP-----ELAAASKLTESRLYDSYHAELRW-----WTG 188  
D 271 YVEPPAEDDRPGRGDWGEGRTRAVQATAMMRLAESGASAAFWNPORTRKACPGCLMRS 330  
QY 189 AFETSEB--IPHSLSVAASDRAVTFGRDRP-----VVANTDRPFRGHDRSVLV 237





PD 09-AUG-2001.  
XX  
XX  
PF 05-FEB-2001; 2001WO-US004098.  
XX  
XX PR 03-FEB-2000; 2000US-00496914.  
PR 27-APR-2000; 2000US-00560875.  
PR 20-JUN-2000; 2000US-0058075.  
PR 19-JUL-2000; 2000US-00620325.  
PR 01-SEP-2000; 2000US-0064936.  
PR 15-SEP-2000; 2000US-00653561.  
PR 20-OCT-2000; 2000US-00693325.  
PR 30-NOV-2000; 2000US-00728422.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang Y, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
PI Ma Y, Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
XX WPI; 2001-476283/51.  
XX  
XX N-PSDB; AAK52402.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful  
PT in diagnosis and gene therapy.  
XX  
XX Claim 20; Page 4333-4335; 6221pp; English.  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM7323-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
CC (AAK52582) and 3666 (AAH80020) are omitted as the relevant pages from the  
CC sequence listing were missing at the time of publication  
XX  
XX Sequence 1176 AA;  
SQ  
Query Match 5.8%; Score 100.5; DB 4; Length 1176;  
Best Local Similarity 22.0%; Pred. No. 1.7;  
Matches 57; Conservative 38; Mismatches 87; Indels 77; Gaps 13;  
QY 86 PPNP--NDPLHLA-SIDFSPADFTVTEGRLRADAIL-----LRTDR 124  
DB 50 PPYPMMNGRLHGHFTSLSKCEFAVGQYRLKGKCLPPFGLHCTGMPKACADKLKREIE 109  
QY 125 LPFAEPPPW-DLVESQRLTYTADTVRIDVADMRPELAA---ASK-----LTESLRLYD 176  
DB 110 L-YGCPDPFPDEEBEETSVKTEDIIIKKAKKKKKAAGSSKYQWGMKMSLGLSD 168  
QY 177 S---SYHAELEFW-----WTGAFETSEGIPIHS-----SLVSAES 207  
DB 169 EELVFESEAEHMLDYFPPLAIDOLKRMGLKVDWRRSFTITDVNPFYDSFVRWQFLTLRER 228  
QY 208 DRVTFGRDPFVAVNTDRRPERGHDSS-----KVLVSTYDNERASLLRGE- 253  
DB 229 NKIKGKRYTIVSPKDGQPCMDHROGEGVGPOEYTLTKLVLEPYPSTKLSGLK--GKN 286  
QY 254 --MLSAVLLDATMAGLATC 270  
DB 287 IFLVATLRLPETMFGQTNC 305

RESULT 11  
ADL13064  
ID ADL13064 standard; protein; 1176 AA.  
XX  
XX AC ADL13064;

XX  
XX 06-MAY-2004 (first entry)  
DT  
XX  
XX DE Human steroid-induced C3A liver cell protein #122.  
XX  
XX KM Hepatotropic; Gene therapy; Wilson disease; liver disorder;  
XX KM steroid therapy; cirrhosis; hepatitis; human; C3A liver cell.  
XX  
XX OS Homo sapiens.  
XX US6673549-B1.  
PN  
XX 06-JAN-2004.  
PD  
XX 12-OCT-2001; 2001US-00976594.  
PF  
XX 12-OCT-2000; 2000US-0240409P.  
PR  
XX (INCY-) INCYTE CORP.  
PA  
XX Furness LM, Buchbinder JL;  
PI  
XX WPI; 2004-068610/07.  
XX  
XX Combination useful for preparing a composition for treating liver  
PT disorders associated with steroid therapy, e.g., cirrhosis or hepatitis,  
PT comprises cDNAs that are differentially expressed in response to steroid  
PT treatment.  
XX  
XX Disclosure; SEQ ID NO 793; 141pp; English.  
XX  
XX The invention relates to a combination comprising cDNAs that are  
CC differentially expressed in response to steroid treatment. Also included  
CC are the following: a high throughput method for using a cDNA to detect  
CC differential expression of nucleic acids in a sample; and a high  
CC throughput method of screening molecules or compounds to identify a  
CC ligand that specifically binds a cDNA. The sample is from a subject with  
CC Wilson disease and comparison of a standard defines a stage of that  
CC disease. The high throughput method of screening molecules or compounds  
CC to identify a ligand that specifically binds a cDNA comprises: combining  
CC the combination with molecules or compounds under conditions to allow  
CC specific binding; and detecting specific binding between each cDNA and at  
CC least one molecule or compound. The molecules or compounds are regulatory  
CC proteins. The combination is useful for preparing a composition for  
CC treating liver disorders associated with steroid therapy, e.g., cirrhosis  
CC or hepatitis. The present sequence represents a human protein which is  
CC differentially expressed in steroid-induced C3A liver cells. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format directly from USPRO  
CC at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
XX  
XX Sequence 1176 AA;  
SQ  
Query Match 5.8%; Score 100.5; DB 8; Length 1176;  
Best Local Similarity 22.0%; Pred. No. 1.7;  
Matches 57; Conservative 38; Mismatches 87; Indels 77; Gaps 13;  
QY 86 PPNP--NDPLHLA-SIDFSPADFTVTEGRLRADAIL-----LRTDR 124  
DB 50 PPYPMMNGRLHGHFTSLSKCEFAVGQYRLKGKCLPPFGLHCTGMPKACADKLKREIE 109  
QY 125 LPFAEPPPW-DLVESQRLTYTADTVRIDVADMRPELAA---ASK-----LTESLRLYD 176  
DB 110 L-YGCPDPFPDEEBEETSVKTEDIIIKKAKKKKKAAGSSKYQWGMKMSLGLSD 168  
QY 177 S---SYHAELEFW-----WTGAFETSEGIPIHS-----SLVSAES 207  
DB 169 EELVFESEAEHMLDYFPPLAIDOLKRMGLKVDWRRSFTITDVNPFYDSFVRWQFLTLRER 228  
QY 208 DRVTFGRDPFVAVNTDRRPERGHDSS-----KVLVSTYDNERASLLRGE- 253  
DB 229 NKIKGKRYTIVSPKDGQPCMDHROGEGVGPOEYTLTKLVLEPYPSTKLSGLK--GKN 286



```
OY 254 --MSAVLDDATMGLATC 270
Db 287 IFLVATLRLPRTMFGQINC 305

RESULT 12
ADFS4681
ID ADFS4681 standard; protein; 897 AA.
XX
XX ADFS4681;
XX
XX 12-FEB-2004 (first entry)
XX
XX Aconitase protein #2.
XX
XX Aconitase; acetic acid resistance; vinegar; acetic acid; enzyme.
XX
XX Gluconacetobacter entanii .
XX
XX JF2003289867-A.
XX
XX 14-OCT-2003.
XX
XX 01-APR-2002; 2002JP-00098589.
XX
XX 01-APR-2002; 2002JP-00098589.
XX
XX (MITS-) MITSUKAN GROUP HONSHA KK.
XX
XX WPI; 2004-038429/04.
XX
XX N-PSDB; ADFS4680.
XX
XX Novel protein having aconitase activity, useful for manufacturing vinegar
XX having high acetic acid concentration.
XX
XX Claim 2; SEQ ID NO 4; 56pp; Japanese.
XX
XX The invention relates to an aconitase protein or a modified protein
XX having one or more amino acid deletions, substitutions or additions and
XX having aconitase activity. The invention also relates to DNA encoding a
XX protein of the invention. The DNA is useful for increasing the acetic
XX acid resistance of microorganisms, preferably acetic acid bacteria
XX belonging to Acetobacter or Gluconacetobacter genus. A microorganism
XX containing DNA and having aconitase activity is useful for manufacturing
XX vinegar. This sequence represents a Gluconacetobacter entanii aconitase
XX protein of the invention.
XX
XX Sequence 897 AA;
XX
XX Query Match 5.8%; Score 99; DB 8; Length 897;
XX Best Local Similarity 24.1%; Pred. No. 1.6;
XX Matches 68; Conservative 31; Mismatches 85; Indels 98; Gaps 15;

OY 46 DRVLYATDHSRGRLIG-----CGAVLDHFRVAAAGTTANVERFPNPNDPLHLASIDF 100
Db 268 DLVLTVTOMLRKKGVGFVFFGPALDHLFVADRA--TINMA-----PEYGATGCF 318
OY 101 SPADFVT-----EGHRLRADAILRLRTDLRPFAL--PPMDLVESQLRTVTYADT 148
Db 319 FPFVDLTLDYLRQGRREHRIKLTAEYLKAGQMRHNASARPVFTDLEMLLETIVS-- 376
OY 149 VRIVVIADDMRPDLAASKLTESLRLYDSYHAELFWMTGAFETSEGIPIHSSLVSAESD 208
Db 377 -----IAGPKRPQDRV-----LKGAIDKAFKEKL--TG-----SLGVPE-----ADKD 412
OY 209 RVTEGRDFPVVANTDRREFGHDSSKVLVSTYDNERASLRGCEMLSAVLDDATMGLA 268
Db 413 KKA-----KVGATNY--EIGHG-----DVIYAIRT 435
OY 269 TCTLTHITELHASRDLVALIGPA-----TPQALVVRGLAP 305
Db 436 SCTWTS-----NPAVLIAAGLVAKKARLGLKPKPVKTSLAP 473
```

```
RESULT 13
AAU48145
ID AAU48145 standard; protein; 459 AA.
XX
XX AAU48145;
XX
XX 27-FEB-2002 (first entry)
XX
XX Propionibacterium acnes immunogenic protein #9041.
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.
XX
XX Propionibacterium acnes.
XX
XX WO200181581-A2.
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US012865.
XX
XX 21-APR-2000; 2000US-0199047P.
XX
XX 02-JUN-2000; 2000US-0208841P.
XX
XX 07-JUL-2000; 2000US-0216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAM, Persing DH, Mitcham UL, Mang SS, Bhactia A;
XX L'maisonmeuve U, Zhang Y, Jen S, Carter D;
XX WPI; 2001-616774/71.
XX
XX N-PSDB; AAS59542.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
XX vaccinating against and diagnosing infections, especially useful for
XX treating acne vulgaris.
XX
XX Example 1; SEQ ID NO 9340; 1069pp; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
XX downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by
XX enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 459 AA;
XX
XX Query Match 5.7%; Score 97.5; DB 4; Length 459;
XX Best Local Similarity 21.2%; Pred. No. 0.85;
XX Matches 73; Conservative 33; Mismatches 129; Indels 109; Gaps 13;

OY 70 FRVAM-AAAGTTANVERFPNPNDPLHLASIDFSPADVFTEGH-----RLRADAILRLRTDR 124
Db 105 YTVAVDAAHGITTGIDAVQARARVLAADSSSPDLVVRGHVLPRLARAGVLERGC-- 162
```

```

QY 125 LRFAPPPDWLVESQLRRTTADTVRI-----DVIADD--MRPELAASKLTES 171
DB 163 -----HTEAAVDLARIAGLEPVGLIGELVDDGMCILRADAIADLARTEG 206
QY 172 LRLYDSSYHAELFWMTGAFETSEGIPIHSSLSVSAESDRVTFGRPFPPVAVNDR----- 224
DB 207 LALTTID---QLAQWRQAHDPGPAVPTQRTVLAASAHLPTRHGGFAITGYRDNCTGVENV 263
QY 225 -----RPEFGHDSKVLVLTSTYDNERASLRLCGEMLS-----AVTL 260
DB 264 LLVGEKGIADDDGGPAWVRVHSECLTGDALGSLRCD---CGDQLAAAOQHVCRRHGAIIL 320
QY 261 DATMAGLATCTTLTHITELHA-----SRDLVAA----- 287
DB 321 LRDEGRATGILINKIAAYKADGGDITVDAQTHGLPVDAREYGAVALILANIGITSVRL 380
QY 288 LIGOPATPOLVVRVGLAPEMEPPATPRRPIDVFFHVRAKDR 331
DB 381 LTNNPAKAEALRDGGL--EVTMSPLETSTREPDERY-LRTKRDR 421

```

## RESULT 14

ABM44664 standard; protein; 459 AA.

ABM44664;

20-OCT-2003 (first entry)

Propionibacterium acnes predicted ORF-encoded polypeptide #9340.

Acne vulgaris; antiseborrheic; dermatological; antibacterial;

immunostimulant; immune response; vaccine.

Propionibacterium acnes.

WO2003033515-A1.

24-APR-2003.

11-OCT-2002; 2002WO-US032727.

15-OCT-2001; 2001US-00978825.

(CORI-) CORIXA CORP.

Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL,

Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D,

Barth B, Valliave-Douglas J;

WPI; 2003-381789/36.

N-P8DB; ACF64471.

Example 1; SEQ ID NO 9340; 1481pp; English.

The invention relates to an isolated polynucleotide (ACF64435-ACF64733)

encoding a Propionibacterium acnes protein. The invention also relates to

polynucleotides encoded by the polynucleotides (ABM35624-ABM64536) and to

immunogenic fragments of P. acnes polypeptides. The invention

additionally encompasses expression vectors and host cells comprising a

polynucleotide of the invention; antibodies against polypeptides of the

invention; fusion proteins comprising a polypeptide of the invention; a

method for stimulating an immune response specific for a P. acnes

polypeptide and an isolated T cell population comprising T cells prepared

via this method; a vaccine composition (comprising P. acnes polypeptides,

polynucleotides, antibodies, fusion proteins, T cell populations, or

antigen-presenting cells that express the polypeptide); a method and kit

for detecting or determining the presence or absence of P. acnes in a

patient; and a method for inhibiting the development of P. acnes in a

CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating acne  
 CC vulgaris, or for stimulating an immune response specific for a P. acnes  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridisation. The vaccine composition is useful for the  
 CC stimulation of an immune response against P. acnes, or for treating acne,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
 CC reading frame) contained within the P. acnes polynucleotides of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from Wipo. [http://pub/published\\_pct\\_sequences](http://pub/published_pct_sequences)

CC Sequence 459 AA;

Query Match 5.7%; Score 97.5; DB 6; Length 459;  
 Best Local Similarity 21.2%; Pred. No. 0.85;  
 Matches 73; Conservative 33; Mismatches 129; Indels 109; Gaps 13;

```

QY 70 FRVAM-AAAGTTAVVERFPNNDPLHLASIDFSPADFTVTEGH-----RLRADAILRRTRD 124
DB 105 YTVAVDAAHGTTTGIDAVQPARKTARVLADPESSTSDLVRRGHVPLABAGVLEVRG-- 162
QY 125 LRFAPPPDWLVESQLRRTTADTVRI-----DVIADD--MRPELAASKLTES 171
DB 163 -----HTEAAVDLARIAGLEPVGLIGELVDDGMCILRADAIADLARTEG 206
QY 172 LRLYDSSYHAELFWMTGAFETSEGIPIHSSLSVSAESDRVTFGRPFPPVAVNDR----- 224
DB 207 LALTTID---QLAQWRQAHDPGPAVPTQRTVLAASAHLPTRHGGFAITGYRDNCTGVENV 263
QY 225 -----RPEFGHDSKVLVLTSTYDNERASLRLCGEMLS-----AVTL 260
DB 264 LLVGEKGIADDDGGPAWVRVHSECLTGDALGSLRCD---CGDQLAAAOQHVCRRHGAIIL 320
QY 261 DATMAGLATCTTLTHITELHA-----SRDLVAA----- 287
DB 321 LRDEGRATGILINKIAAYKADGGDITVDAQTHGLPVDAREYGAVALILANIGITSVRL 380
QY 288 LIGOPATPOLVVRVGLAPEMEPPATPRRPIDVFFHVRAKDR 331
DB 381 LTNNPAKAEALRDGGL--EVTMSPLETSTREPDERY-LRTKRDR 421

```

## RESULT 15

ABP69805 standard; protein; 585 AA.

ABP69805;

20-JAN-2003 (first entry)

Human polypeptide SEQ ID NO 1852.

Human; genome mapping; gene therapy; food supplement; virus; fungus;

cell-proliferative disorder; neurodegenerative disease; bacterial;

Parkinson's disease; Alzheimer's disease; autoimmune disease; disease;

multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;

arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;

antiparkinsonian; antidiabetic; immunosuppressive; dermatological;

haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;

antiarthritic.

Homo sapiens.

WO200270539-A2.

12-SEP-2002.

05-MAR-2002; 2002WO-US005095.

05-MAR-2001; 2001US-00799451.



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OW protein - protein search, using sw model

Run on: March 23, 2006, 05:11:18 ; Search time 13.1689 Seconds  
(without alignments)  
2418.401 Million cell updates/sec

Title: US-10-617-038-25  
1718

Sequence: 1 MPDMMVTVDVKSAVQLACR.....PATPRRPIDEVFHVRKDKHR 331

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1718	100.0	331	2	G70942
2	912	53.1	344	2	F70922
3	417	24.3	332	2	D70645
4	295	17.2	330	2	T36944
5	252	14.7	335	2	T37044
6	171	10.0	309	2	T37033
7	143	8.3	725	2	B70741
8	121	7.0	372	2	B70661
9	109.5	6.4	213	2	A84250
10	109	6.3	409	2	C87319
11	108	6.3	721	2	A87128
12	104.5	6.1	382	2	AG3161
13	102.5	6.0	395	2	T34908
14	99	5.8	434	2	F75425
15	98	5.7	373	2	T34743
16	96.5	5.6	481	2	C64945
17	96	5.6	2082	2	T37056
18	95.5	5.6	481	2	A89947
19	95.5	5.6	481	2	B85795
20	94	5.5	602	2	T50974
21	94	5.5	1111	2	T38407
22	93.5	5.4	201	2	E95072
23	93.5	5.4	765	2	H84247
24	92.5	5.4	547	2	G71307
25	92.5	5.4	1131	2	T15617
26	92	5.4	1186	2	T19334
27	91.5	5.3	559	2	B84213
28	91.5	5.3	750	2	H70711
29	91	5.3	516	1	S72937

30	91	5.3	517	2	G87032	conserved hypotet
31	91	5.3	635	2	F70874	probable membrane
32	90.5	5.3	383	2	A95871	probable glycosyl
33	90.5	5.3	421	2	T35205	citrate synthase-1
34	90.5	5.3	457	2	G87003	probable oxidoredu
35	90	5.2	881	2	F84404	valyl-tRNA synthet
36	89.5	5.2	201	2	B97940	nitroreductase (EC
37	89.5	5.2	1173	2	T25539	hypothetical prote
38	89.5	5.2	1615	2	B49502	protein-tyrosine-p
39	89.5	5.2	1767	2	A49502	protein-tyrosine-p
40	89.5	5.2	1802	2	H88444	protein C26E.12 [
41	89	5.2	283	2	C84321	hypothetical prote
42	89	5.2	448	2	G70977	hypothetical prote
43	89	5.2	1143	2	A69465	DNA-directed DNA p
44	89	5.2	1150	2	S58775	myPI protein - smu
45	88.5	5.2	1059	2	T21891	hypothetical prote

ALIGNMENTS

RESULT 1

G70942  
Hypothetical protein Rv2032 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
C:Accession: G70942  
R:Coile, S.T.; Brosch, R.; Parthill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: G70942  
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-331 <COL>  
A:Cross-references: UNIPROT:O53476; UNIPARC:UPI00001652C1; GB:AL021899; GB:AL123456; NID  
A:Experimental source: strain H37RV  
C:Gene: Rv2032  
C:Superfamily: Mycobacterium hypothetical protein Rv3131

Query Match	Similarity	Score	Length	DB 2;	Length	331;
Best local	100.0%	1718	331			
Matches	331;	Conservative	0;	Mismatches	0;	Indels
0;						Gaps
0;						
Qy	1	MPDMMVTVDVKS	AVQLACRAPSILHNSQPMWIMAEHTVALFLDKRVLVA	TDHSGREAL	60	
Db	1	MPDMMVTVDVKS	AVQLACRAPSILHNSQPMWIMAEHTVALFLDKRVLVA	TDHSGREAL	60	
Qy	61	LGCAVLDHFRVNA	AAAGTTANVERFPNPNDPLHLASIDSPADPVTGHRRLRDA	ILLR	120	
Db	61	LGCAVLDHFRVNA	AAAGTTANVERFPNPNDPLHLASIDSPADPVTGHRRLRDA	ILLR	120	
Qy	121	RTDRLPFAEPD	WDLVSQRLTATYADTVRIDYADDMRPELAASKLTESLR	LYDSYH	180	
Db	121	RTDRLPFAEPD	WDLVSQRLTATYADTVRIDYADDMRPELAASKLTESLR	LYDSYH	180	
Qy	121	RTDRLPFAEPD	WDLVSQRLTATYADTVRIDYADDMRPELAASKLTESLR	LYDSYH	180	
Db	121	RTDRLPFAEPD	WDLVSQRLTATYADTVRIDYADDMRPELAASKLTESLR	LYDSYH	180	
Qy	181	AELFMWTCAPF	ETSGIRHSSLSVSAESDRVTFGRDPFVANTDRPFGHRSKVL	VLST	240	
Db	181	AELFMWTCAPF	ETSGIRHSSLSVSAESDRVTFGRDPFVANTDRPFGHRSKVL	VLST	240	
Qy	241	YDNERASILRC	GENLSAVLLDATMAAGLATCTLTHITELHRSRLVAALIGQ	PATPOLVR	300	
Db	241	YDNERASILRC	GENLSAVLLDATMAAGLATCTLTHITELHRSRLVAALIGQ	PATPOLVR	300	
Qy	301	VGLAPEMEEP	PATPRRPIDEVFHVRAKDKHR	331		
Db	301	VGLAPEMEEP	PATPRRPIDEVFHVRAKDKHR	331		





Db 596 TNERV-----SASSALATVYVDGATLTDPARGSGMAQAVIYAQQHGLAVQMSPIFLYAR 651  
Qy 276 -----TELHA-----SRDLVVALIGOPATPOLVAVGLAPMEBPPATPRRP 318  
Db 652 GRHLDGASPHFAQHLQDPRFLVAP--GKSGHEVLIRLPHAP-----PFSVCSRR- 704  
Qy 319 IDEVFHVRAKDHR 331  
Db 705 --RVRHAIPEPHR 715

RESULT 8  
B70661  
hypothetical protein RV2337c - Mycobacterium tuberculosis (strain H37RV)  
C/Species: Mycobacterium tuberculosis  
C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
C/Accession: B70661  
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A/Reference number: A70500; PMID:98295987; PMID:9634230  
A/Accession: B70661  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-372 <COL>  
A/Cross-references: UNIPROT:P95233; UNIPARC:UPI00000CCBF7; GB:Z83860; GB:AL123456; NID:9  
A/Experimental source: strain H37RV  
C/Genetics:  
A/Gene: RV2337c  
C/Superfamily: Mycobacterium hypothetical protein RV3131

Query Match 7.0%; Score 121; DB 2; Length 372;  
Best Local Similarity 23.9%; Pred. No. 0.0095;  
Matches 88; Conservative 35; Mismatches 121; Indels 124; Gaps 17;  
Db 13 SAVOLACRAPSLAH-----SQPMRWIAEDHTVALFLDKDRLVATDHS 55  
Qy 20 SLVEAALALAPESADNRREVQLEHAGRVRLMGDQTM-----SADENR 61  
Db 56 GREALLCGAVLDHFRVMAAAGTTANVERPNDPLHLASIDFSPADFTGHRRLAD 115  
Qy 62 RIMSLVAIGAAVENVKRALGKETKVCWPDGSPGLVAEID-----VDRLLPQTRVD 115  
Db 116 ----AILLRTRD-RLPFAEPDMDLVESQRTVTATDVRIDVI-----ADNRPELAA 164  
Qy 116 PISGALIRRRKTRNVRFRGP--LSQGL-GLASAEATGIDGQLQWMPSPETRKQILR 171  
Db 165 ASKLTESLRVDSSTYHAEF-----WMTGAFETSEGIPIHSSLVSAESDRTVGRDFP 217  
Qy 172 LVRLAEFRPRSRRLHELFSAVAFDGLMTAS--SDGLPFGSLVEVAM----- 218  
Db 218 VVANTDRPRP-CHDRSKVTLVLTVDNERASLACGEMLSAVLIDATMAGLATCTLR-HI 275  
Qy 219 -----NRPMFRGILHRVVL-----RLRTVGCHHNLG---RAATVLPRLAPHV 259  
Db 276 TELHASRDIVA-----ALIGAPTPQALVAVGLAPMEBPPATPRRP 319  
Qy 260 GALTTSIDLDSGALTAGAVFERILKRTTLTGAELQPPAASVLSLPACENAP----- 312  
Db 320 DEVFHVR 327  
Qy 313 ----HVR 316

RESULT 9  
A84250  
NADH oxidase [imported] - Halobacterium sp. NRC-1  
C/Species: Halobacterium sp. NRC-1  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C/Accession: A84250

R/Ng, W.V.; Kennedy, S.P.; Mahites, G.G.; Bergiste, B.; Pan, M.; Shukla, H.D.; Laeky, S.; Lethausser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor Jung, K.H.; Alam, M.; Freilich, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A/Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li, A/Title: Genome sequence of Halobacterium species NRC-1.  
A/Reference number: A84160; PMID:20504483; PMID:11016950  
A/Accession: A84250  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-213 <STO>  
A/Cross-references: UNIPROT:Q9HQZ4; UNIPARC:UPI0000063780; GB:AE004437; NID:g10580495; P  
C/Genetics:  
A/Gene: noxC  
C/Superfamily: nitroreductase

Query Match 6.4%; Score 109.5; DB 2; Length 213;  
Best Local Similarity 20.1%; Pred. No. 0.044;  
Matches 66; Conservative 27; Mismatches 83; Indels 153; Gaps 14;  
Db 2 PDMVTVDVKSVAVOLACRAPSLHNSQPMRWIAEDHTVALFLDKR-----VLVATDHS 55  
Qy 28 PDMDIDDDTLAELIRDAITLAPSSYNLQPMFEFVQ-----DDRLAELAIAYDQH- 79  
Db 56 GREALLCGAVLDHFRVMAAAGTTANVERPNDPLHLASIDFSPADFTGHRRLAD 115  
Qy 80 ----YDAG-----TALIVAGHT-----DP-----KTADRVF----- 103  
Db 116 AILLRTRDRLPFAEPDMDLVESQRTVTATDVRIDVIADNRPELAAASKLTESRLY 175  
Qy 104 -----EM-----VDGR-----F 111  
Db 176 DSSYHAEFLFWMTGAFETSEGIPIHSSLVSAESDRTVGRDFPVANTDRRPERGHRSKV 235  
Qy 112 DADTGAEL-----KSQTVASYESDQA--GRUYAL----- 138  
Db 236 LVSTVDNERASLIRCGEMLSAVLIDATMAGLATCTTLTELHASRDVLAALIGPATP 235  
Qy 139 -----RVASL-----AAQNLILSHARGLATPMSGP-----DEAAAPAGLPADT 180  
Db 296 QALVAVGLAPMEBPPATPRPIDEVFH 324  
Qy 181 VPVVLIVGPSGGEPPRLPRSSVDEVH 209

RESULT 10  
C87319  
hypothetical protein CC0565 [imported] - Caulobacter crescentus  
C/Species: Caulobacter crescentus  
C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
C/Accession: C87319  
R/Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.; B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Dukkin, A.S.; Gilm, M.L.; Haft, D.H.; Kolon, n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A/Title: Complete Genome Sequence of Caulobacter crescentus.  
A/Reference number: A87249; PMID:21173698; PMID:11259647  
A/Accession: C87319  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-409 <STO>  
A/Cross-references: UNIPROT:Q9AAN2; UNIPARC:UPI00000C70B6; GB:AE005673; NID:g13421759; P  
C/Genetics:  
A/Gene: CC0565

Query Match 6.3%; Score 109; DB 2; Length 409;  
Best Local Similarity 23.2%; Pred. No. 0.12;  
Matches 78; Conservative 34; Mismatches 128; Indels 96; Gaps 14;  
Qy 17 IACRAPSILHNSQPMRWIAEDHTVALFLDKRVLYAT-----DHSGREALLCGC-----AVL 67  
Db 35 LSAVAFELKKGGDGR-----AHALSRSVGYATVRLQTSRGKALVGRGPLASAE 86



```

Qy 68 DFRVMAAAGTTA-NVERPNDNDPLHASTIDFSPADFTVTEGRLRADAILLRTDRLP 126
Db 87 DGLVLTFTALSGAVNLDLDPDHP-----GPRL-----IVLPKWTAMP 123
Qy 127 FAEPDMDLVESQRLTATYADTVRIDVIAADMREPILAAASKUTE--SLRLVSSVHAELF 184
Db 124 DPNRKGW-----ALATRLVNDASALVLPADLRKCLTLRAPPEVGPRILTRPD--GOLF 176
Qy 185 -----WMTGAFTSEGIPIHSSLSVSAESDRVTFGRDPFVVANTDRRPEFG 229
Db 177 GLQPSVRAPRSLGKGMTPVLKDQDG-----RMVLAMHTESQTVYLAEPDLNTATLKTLE 232
Qy 230 HDRSKVTLVSTYDNERASLRCGEMLSAVLLDATMAGLATCTVTHITELHASRDVLAALI 289
Db 233 GARTVALLDLIDHAE-----TPVFEDATLHGFA-----PARSLRLIL 271
Qy 290 GQAPTPQALVRV-----GLAPEMEPPPTPRPRI 319
Db 272 EPLVGATVLVALAVALAGLQAGVRFGLKERRAV 307

```

## RESULT 11

```

A87128
conserved hypothetical protein ML1751 [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C:Accession: A87128
R:Coile, S.T.; Eigmeier, K.; Pakhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthey, S.; Feltewell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.W.
Nature 409, 1007-1011, 2001
A:Authors: Ruter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: A87128
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-721 <STO>
A:Cross-references: UNIPROT:Q9CBP5; UNIPARC:UP100000C6DE0; GB:AL450380; NID:gl3093490; F
C:Genetics:
A:Gene: ML1751

```

```

Query Match 6.3%; Score 108; DB 2; Length 721;
Best Local Similarity 25.1%; Pred. No. 0.32;
Matches 89; Conservative 34; Mismatches 103; Indels 128; Gaps 18;

Qy 10 VIKSNVQLACRAPSLSHNPWR-----WIAEDHTVALFLDKRVLATDHSGRBA 59
Db 379 VIATA--AIRAPSGNLQPMHTIDAGPDALVIRLAPHTDSL-----DVGFRGS 424
Qy 60 LLGGAVLDHFRVMAAAGT-----TANVERFP-----NPN-----DPLHL-- 95
Db 425 AVANGALLFNARVMAAAGVGLCPVSLAENVGAPRLVLSLGDGKNPDLAAVYEMLVRE 484
Qy 96 -----ASIDFSPADFTVTEGRL-----RAD-----AILRTDRLPFAEPDWM 133
Db 485 TRNHNKPTWVLTETIDVLAHAAVEGARHLHLTRDDIVRAATILAAADSRVYTRPLH 544
Qy 134 DLVESQLATTYTA--DTRIVDIADMR-PELAAASKU--TESLRLYSSVTHAELFWTGS 188
Db 545 EEMISELFWPGSPDPT-GIDVRSLEMDWHQLAADVLRRTDVMELAQ-----NWVG 596
Qy 189 AFETSEGIPIHSSLSVSAESDRVTFGRDPFVVANTDRRPEFGDR--SKVLVSTYDNE 244
Db 597 A-----ALGQD-----THDRVASSALAVLSVPGHA 622
Qy 245 RASLRCCEMLSAVLLDATMAGLATCTLT-----HITELHASRDVLAALIGQ 291
Db 623 LADYARSGSAVEALIVVAHQRYGFAVQPISPVFLVHSDVDELHAASAQFADELGQ 676

```

```

conserved hypothetical protein Atu5016 [imported] - Agrobacterium tumefaciens (strain C5
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AG3161
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, L
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E
ser, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AG3161
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-382 <KOR>
A:Cross-references: UNIPROT:Q8UKT3; UNIPARC:UP100000D1489; GB:AB008687; PIDN:AA145709.1;
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu5016
A:Genome: plasmid
C:Superfamily: Rhizobium plasmid pNGR234a yacP protein

```

Query Match 6.1%; Score 104.5; DB 2; Length 382;

Best Local Similarity 21.5%; Pred. No. 0.27;

Matches 73; Conservative 36; Mismatches 114; Indels 117; Gaps 15;

```

Qy 69 HFRVMAAAGTTANVERPNDNDPLHASTIDFSPA--DFVTEGRLRADAILLRTDRLP 126
Db 48 HFTACSLWVDGSL-----VHLEDVLHATDITPTHTLTIARDVLTARRRIA 98
Qy 127 FAEPDMDLVESQRLT-----TVTADTVRIDVIAAD----- 157
Db 99 -AGPPDWTLSADGRLTRQSTSEIASTVGAEETAAIRRVVTDPEGEDEVDGSENLPG 157
Qy 158 -----MRPELAASKUTE-SLELYSSY--HAEFLWTAFTSE 194
Db 158 VDLEAIDAVLARSEAIASATRPGRAGSRAATEKDPVYDLDWDEEARLDEMGRVLRQAO 217
Qy 195 GIP-----HSSLSVSAESDRVTFG-RDPPV 219
Db 218 ELPAVLDQAIYALDAMNELSVQHPWIGRLAASILHNAAGVTSGLHAAIYGLKTFPV- 276
Qy 220 ANTDREPEFGHRSKVLVSTYDNERASLRCGE---MLSAVLLDATMAGLATCT-LTH 274
Db 277 ---DRRH--RDRSRLLAIHGFIAAEIGLKHDRLLIARTWVERKLDGRTSSKLP 331
Qy 275 ITELHASRDVLAALIGQPA-----TPQALVRVGLAPEMER 309
Db 332 LVLELVMAKPLVSA--GNVAKTLDVTPQARRIVLELGLRE 369

```

## RESULT 13

```

T34908
redv protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004
C:Accession: T34908
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL data library, January 1998
A:Reference number: Z21558
A:Accession: T34908
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-395 <OLI>
A:Cross-references: UNIPROT:O54145; UNIPARC:UP100000DAB97; EMBL:AL021409; PIDN:CAA16173.1
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: redV; SCOREDB:SC3F7.02c

```

Query Match 6.0%; Score 102.5; DB 2; Length 395;

Best Local Similarity 21.8%; Pred. No. 0.42;

Matches 86; Conservative 43; Mismatches 166; Indels 99; Gaps 16;

## RESULT 12

AG3161





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 04:56:43 ; Search time 84.6574 Seconds  
(without alignments)  
2758.529 Million cell updates/sec

Title: US-10-617-038-25

Perfect score: 1718  
Sequence: 1 MPDMMVTVDVKSAYQLACR.....PATPRRPIDVEFHVHAKDHR 331

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_05.80:\*  
1: uniprot\_prot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1718	100.0	331	053476_MYCTU	053476 mycobacteri
2	1718	100.0	333	07D712_MYCTU	07D712 mycobacteri
3	1708	99.4	331	07VETO_MYCBO	07VETO mycobacteri
4	915	53.3	348	07D623_MYCTU	07D623 mycobacteri
5	912	53.1	344	005800_MYCTU	005800 mycobacteri
6	904	52.6	344	07X774_MYCBO	07X774 mycobacteri
7	847.5	49.3	333	07X260_MYCPA	07X260 mycobacteri
8	629	36.6	333	05YVW2_NOCFA	05YVW2 nocardia fa
9	477.5	27.8	335	05YVW8_NOCFA	05YVW8 nocardia fa
10	434.5	25.3	324	073UJ3_MYCPA	073UJ3 mycobacteri
11	417.5	24.3	342	073257_MYCPA	073257 mycobacteri
12	417	24.3	332	P95195_MYCTU	P95195 mycobacteri
13	417	24.3	344	07X770_MYCBO	07X770 mycobacteri
14	417	24.3	344	07D627_MYCTU	07D627 mycobacteri
15	334.5	19.5	343	063H03_MYCBO	063H03 mycobacteri
16	295	17.2	330	09R103_STRCO	09R103 streptomyce
17	266.5	15.5	375	04NVM0_PDELT	04NVM0 anaeromyxob
18	252	14.7	309	09R133_STRCO	09R133 streptomyce
19	171	10.0	335	09R144_STRCO	09R144 streptomyce
20	169	9.8	382	04IXJ3_AZOVI	04IXJ3 azotobacter
21	163	9.5	382	05LTM6_SILPO	05LTM6 silicobacter
22	156.5	9.1	393	08TND4_METAC	08TND4 methanobarc
23	155.5	9.1	330	07WVPS_PORGI	07WVPS porphyromon
24	147	8.6	352	05W242_SERMA	05W242 serratia ma
25	143	8.3	715	Y11355_MYCTU	Y11355 mycobacteri
26	143	8.3	715	07U074_MYCBO	07U074 mycobacteri
27	140.5	8.2	989	07ZM05_DESJA	07ZM05 desulfocylr
28	125	7.3	366	08BRM5_BRAJA	08BRM5 bradyrhizob
29	124	7.2	357	05NKK0_AZOSE	05NKK0 azarconus sp
30	121	7.0	372	07D7A9_MYCCTU	07D7A9 mycobacteri
31	121	7.0	372	07TYS8_MYCBO	07TYS8 mycobacteri

32	121	7.0	372	2	P95233_MYCTU	P95233 mycobacteri
33	120	7.0	449	1	COFE_MYCPA	073UJ3 mycobacteri
34	118.5	6.9	713	2	05YVX2_NOCFA	05YVX2 nocardia fa
35	113	6.6	589	2	04NLB3_MITCC	04NLB3 arthrobacter
36	109.5	6.4	213	2	09H024_HALSA	09H024 halobacteri
37	109	6.3	409	2	09AAN2_CAUCR	09AAN2 caulobacteri
38	108	6.3	721	2	09CBP5_MYCLB	09CBP5 mycobacteri
39	106.5	6.2	413	2	09X5U2_STRLA	09X5U2 streptomyce
40	104.5	6.1	382	2	08UKT3_AGRIS	08UKT3 agrobacteri
41	104	6.1	380	2	04INB9_BURK	04INB9 burkholderi
42	103	6.0	528	2	06G1C8_BAROU	06G1C8 bartonella
43	102.5	6.0	395	2	054145_STRCO	054145 streptomyce
44	102	5.9	441	2	08ZCH4_STRAM	08ZCH4 streptomyce
45	101.5	5.9	306	2	04T081_9EPHN	04T081 erythrobact

## ALIGNMENTS

RESULT 1						
ID	053476_MYCTU	PRELIMINARY;	PRT;	331	AA.	
AC	053476-					
DT	01-JUN-1998 (TREMBlrel. 06, Created)					
DT	01-MAR-2002 (TREMBlrel. 20, Last sequence update)					
DT	01-JUN-2003 (TREMBlrel. 24, Last annotation update)					
DE	Hypothetical protein acg.					
GN	Name=acg; OrderedLocustNames=Rv2032;					
OS	Mycobacterium tuberculosis.					
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;					
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium;					
OX	NCB1_TaxID=1773;					
RN	[1]					
RP	NUCLEOTIDE SEQUENCE.					
RC	STRAIN=H37Rv;					
RX	MEDLINE=96295987; PubMed=9634230; DOI=10.1038/31159;					
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churche C.M.,					
RA	Harris D.E., Gordon R., Eigmeier K., Gae S., Barry C.E. III,					
RA	Tekaia F., Badcock K., Baeham D., Brown D., Chillingworth T.,					
RA	Connor R., Davies R.M., Devlin K., Krogan A., McLean S., Moule S.,					
RA	Holroyd S., Hornsby T., Jørgen K., Krogan A., McLean S., Moule S.,					
RA	Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,					
RA	Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,					
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;					
RT	"Deciphering the biology of Mycobacterium tuberculosis from the					
RT	complete genome sequence."					
RL	Nature 393:537-544 (1998).					
DR	EMBL; BX842578; CAA17246.1; -; Genomic_DNA.					
DR	PIR; G70942; G70942.					
DR	TubercuList; Rv2032; -.					
KW	Complete proteome; Hypothetical protein.					
SQ	SEQUENCE 331 Aa; 36559 Mw; 8FF2A825CE90C6B CRC64;					
Query Match						
Best Local Similarity 100.0%; Pred. No. 1e-133;						
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
QY	1	MPDMMVTVDVKSAYQLACRA	PSLHNSOPKRWIADH	VALFLDKDRVLYATD	HSGRAL	60
DB	1	MPDMMVTVDVKSAYQLACRA	PSLHNSOPKRWIADH	VALFLDKDRVLYATD	HSGRAL	60
QY	61	LCGCAVLDPFVMAAAGTTA	VRFPNPDLHLASID	FPADFTVGGHRLAD	AILLR	120
DB	61	LCGCAVLDPFVMAAAGTTA	VRFPNPDLHLASID	FPADFTVGGHRLAD	AILLR	120
QY	121	RTDRLPFAEPDMDLVESQ	LRTTYADTVRIDVIA	DDMRPELAASKLT	ESLRVDSYH	180
DB	121	RTDRLPFAEPDMDLVESQ	LRTTYADTVRIDVIA	DDMRPELAASKLT	ESLRVDSYH	180
QY	181	ABLFMTWTCAGFTSGCI	PHSSIVSAAESDRT	VFSGDFPVVANTDR	RPERGHRSKVLV	240
DB	181	ABLFMTWTCAGFTSGCI	PHSSIVSAAESDRT	VFSGDFPVVANTDR	RPERGHRSKVLV	240

QY 241 YDNERASLLRCGEMTSAVLLDATTMAGLATCTTTHITELHASRDVLAALIGOPATPOLVLR 300  
DB 241 YDNERASLLRCGEMTSAVLLDATTMAGLATCTTTHITELHASRDVLAALIGOPATPOLVLR 300  
QY 301 VGLAPEMEBPPATPRRPIDEVFHVRAKDR 331  
DB 301 VGLAPEMEBPPATPRRPIDEVFHVRAKDR 331

## RESULT 2

Q7D7L2 MYCTU  
ID Q7D7L2 MYCTU PRELIMINARY; PRT; 333 AA.

AC Q7D7L2;  
DT 05-JUL-2004 (TREMBLrel. 27, Created)  
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
DE Hypothetical protein.  
GN OrderedLocustNames=MT2091;  
OS Corynebacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;  
OC Mycobacterium tuberculosis complex.  
OX NCBI\_TaxId=1773;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=CDC 1551 / Oshkosh;  
RX MEDLINE=22206494; PubMed=12218036;  
RX DOI=10.1128/JB.184.19.5479-5490.2002;  
RA Fleischman R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
Peterson J.D., DeBoy R.T., Dodson R.W., Gwinn M.L., Haft D.H.,  
Hickey E.K., Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D.,  
Salzberg S.L., Salzberg A., Uterback T.R., Weidman J.F., Khouri H.M.,  
Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,  
Fraser C.M.;  
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RT J. Bacteriol. 184:5479-5490 (2002).  
DR EMBL; AE000516; AAK46370.1; -; Genomic\_DNA.  
DR TIGR; MT2091; -;  
KW Hypothetical protein.  
SQ SEQUENCE 333 AA; 36803 MW; D873842445EAD5ED CRC64;

Query Match 100.0%; Score 1718; DB 2; Length 333;  
Best Local Similarity 100.0%; Pred. No. 1e-133;  
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPDWTMTDVIKSAVQLACRAPSLHNSQPMRWIAEDHTVALFLDKDRVLYATDHSGREAL 60  
DB 3 MPDWTMTDVIKSAVQLACRAPSLHNSQPMRWIAEDHTVALFLDKDRVLYATDHSGREAL 62  
QY 61 LGGCAVLDFHRVMAAAGTTANVERFPNPNDPLHLASIDFSPADPVTGHRRLADALLR 120  
DB 63 LGGCAVLDFHRVMAAAGTTANVERFPNPNDPLHLASIDFSPADPVTGHRRLADALLR 122  
QY 121 RTDRLPAEPDPMDLVESQRLRTVTADTVRIDVIADDMRPELAAASKLTESLRLYDSSTY 180  
DB 123 RTDRLPAEPDPMDLVESQRLRTVTADTVRIDVIADDMRPELAAASKLTESLRLYDSSTY 182  
QY 181 AELFWMTGAFETSEGIPIHSSLSVAESDRVTFGRDFPVVANTDRRPFEGHDSKVLVLT 240  
DB 183 AELFWMTGAFETSEGIPIHSSLSVAESDRVTFGRDFPVVANTDRRPFEGHDSKVLVLT 242  
QY 241 YDNERASLLRCGEMTSAVLLDATTMAGLATCTTTHITELHASRDVLAALIGOPATPOLVLR 300  
DB 243 YDNERASLLRCGEMTSAVLLDATTMAGLATCTTTHITELHASRDVLAALIGOPATPOLVLR 302  
QY 301 VGLAPEMEBPPATPRRPIDEVFHVRAKDR 331  
DB 303 VGLAPEMEBPPATPRRPIDEVFHVRAKDR 333

## RESULT 3

Q7VE70 MYCBO  
ID Q7VE70 MYCBO PRELIMINARY; PRT; 331 AA.

AC Q7VE70;  
DT 01-OCT-2003 (TREMBLrel. 25, Created)  
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Hypothetical protein MD2058.  
GN OrderedLocustNames=MD2058;  
OS Mycobacterium bovis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;  
OC Mycobacterium tuberculosis complex.  
OX NCBI\_TaxId=1765;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=AF2122/97;  
RX MEDLINE=23709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;  
RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,  
Pryor M., Duchoy S., Grondin S., Lacroix C., Montempe C., Simon S.,  
Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
RA "The complete genome sequence of Mycobacterium bovis.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).  
DR EMBL; BX248341; CAD96911.1; -; Genomic\_DNA.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 331 AA; 36575 MW; BFF2A825CEBBAAB CRC64;

Query Match 99.4%; Score 1708; DB 2; Length 331;  
Best Local Similarity 99.7%; Pred. No. 6.9e-133;  
Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPDWTMTDVIKSAVQLACRAPSLHNSQPMRWIAEDHTVALFLDKDRVLYATDHSGREAL 60  
DB 1 MPDWTMTDVIKSAVQLACRAPSLHNSQPMRWIAEDHTVALFLDKDRVLYATDHSGREAL 60  
QY 61 LGGCAVLDFHRVMAAAGTTANVERFPNPNDPLHLASIDFSPADPVTGHRRLADALLR 120  
DB 61 LGGCAVLDFHRVMAAAGTTANVERFPNPNDPLHLASIDFSPADPVTGHRRLADALLR 120  
QY 121 RTDRLPAEPDPMDLVESQRLRTVTADTVRIDVIADDMRPELAAASKLTESLRLYDSSTY 180  
DB 121 RTDRLPAEPDPMDLVESQRLRTVTADTVRIDVIADDMRPELAAASKLTESLRLYDSSTY 180  
QY 181 AELFWMTGAFETSEGIPIHSSLSVAESDRVTFGRDFPVVANTDRRPFEGHDSKVLVLT 240  
DB 181 AELFWMTGAFETSEGIPIHSSLSVAESDRVTFGRDFPVVANTDRRPFEGHDSKVLVLT 240  
QY 241 YDNERASLLRCGEMTSAVLLDATTMAGLATCTTTHITELHASRDVLAALIGOPATPOLVLR 300  
DB 241 YDNERASLLRCGEMTSAVLLDATTMAGLATCTTTHITELHASRDVLAALIGOPATPOLVLR 300  
QY 301 VGLAPEMEBPPATPRRPIDEVFHVRAKDR 331  
DB 301 VGLAPEMEBPPATPRRPIDEVFHVRAKDR 331

## RESULT 4

Q7D629 MYCTU  
ID Q7D629 MYCTU PRELIMINARY; PRT; 348 AA.

AC Q7D629;  
DT 05-JUL-2004 (TREMBLrel. 27, Created)  
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
DE Hypothetical protein.  
GN OrderedLocustNames=MT3212;  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;  
OC Mycobacterium tuberculosis complex.  
OX NCBI\_TaxId=1773;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=CDC 1551 / Oshkosh;

```

RX MEDLINE=22206494; PubMed=12218036;
RX DOI=10.1128/JB.184.19.5479-5490.2002;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.D., DeBoy R.T., Dodson R.J., Gwinn M.L., Haft D.H.,
RA Hickey E.K., Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D.,
RA Salzberg S.L., Delcher A., Uterback T.R., Weidman J.F., Khouri H.M.,
RA Gill J., Mikita A., Bishai W., Jacobs W.R. Jr., Venter J.C.,
RA Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains ";
RL J. Bacteriol. 184:5479-5490(2002).
DR EMBL; AE000516; AAK47552.1; -; Genomic_DNA.
DR TIGR; MT3212; -.
KW Hypothetical protein.
SQ SEQUENCE 348 AA; 38976 MW; AE78834535008BDA CRC64;

Query Match 53.3%; Score 915; DB 2; Length 348;
Best Local Similarity 56.7%; Pred. No. 3.3e-67;
Matches 182; Conservative 39; Mismatches 96; Indels 4; Gaps 2;

QY 10 VIKSAVOLACRAPSLHNSQPMRWIAE--DH-TVALFLDKDRVLYATDHSGREALLGCGA 65
D 5 VLNKAVALACRAPSVHNSQPMRWVAESGSEHTVHLFVNHRRTVPATDHSGRQAIIISGCA 64
QY 66 VLDHFRVMAAAGTTANVERFPNPNDLHLASIDFSPADFTVTEGHRRLADAILRRDRL 125
D 65 VLDHRLRIMTAHQAQNTIRFPQNPQDLATVEFSPIDHTVAGQRRAQAILQRDRL 124
QY 126 PFAEPPMDLVESQRTTVDIVRIADNMRPELAASKLTESLRLYDSYHAELFW 185
D 125 PFDSPMTWHLFEPLRDAVDKDVAMLDVSDQRTRLVVASQLSEVLRDDPYHAELFW 184
QY 186 WTGAFTSEGIPIHSLVSAESDRVTFGDRPPVANTDRPEFGHDSKVLVYSTYNER 245
D 185 WTSPPVLAHGVPPPTLASDARLRVLDGRDPVRSYQNRRAELADDSKVLVSTPSPDR 244
QY 246 ASILRCGEMTSAVLIDATAGLATCTTLHTLTHLASRDLVAALIGPATPOALVRVGLAP 305
D 245 ADALRCGEVSTILLECTMAGMATCTTLHTLIESDSRDIVGLTRQGEFQALIRVGIAF 304
QY 306 EMEPPPATPRRPIDEVFHV 326
D 305 PLAAPVAPTPRRPLDVSQIR 325

RESULT 5
O05800 MYCTU PRELIMINARY; PRT; 344 AA.
AC O05800;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=RV3127;
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium tuberculosis complex.
OX NCBI_TaxID=1773;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=H37Rv;
RC MEDLINE=98295987; PubMed=9634230; DOI=10.1038/31159;
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C.M.,
RA Harris D.E., Gordon S.V., Eigemeier K., Gao S., Barry C.E. III,
RA Tekaia F., Badcock K., Baaham D., Brown D., Chillingworth T.,
RA Connor R., Davies R.M., Devlin K., Fellwell T., Gentles S., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S.,
RA Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,
RA Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sutcliffe J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
```

```

RL Nature 393:537-544(1998).
DR EMBL; BX842582; CAB08361.1; -; Genomic_DNA.
DR PIR; F70922; F70922.
DR Tuberculist; RV3127; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 344 AA; 38520 MW; 9DACAA3495F03770 CRC64;

Query Match 53.1%; Score 912; DB 2; Length 344;
Best Local Similarity 56.4%; Pred. No. 5.8e-67;
Matches 181; Conservative 40; Mismatches 96; Indels 4; Gaps 2;

QY 10 VIKSAVOLACRAPSLHNSQPMRWIAE--DH-TVALFLDKDRVLYATDHSGREALLGCGA 65
D 1 MLNKAVALACRAPSVHNSQPMRWVAESGSEHTVHLFVNHRRTVPATDHSGRQAIIISGCA 60
QY 66 VLDHFRVMAAAGTTANVERFPNPNDLHLASIDFSPADFTVTEGHRRLADAILRRDRL 125
D 61 VLDHRLRIMTAHQAQNTIRFPQNPQDLATVEFSPIDHTVAGQRRAQAILQRDRL 120
QY 126 PFAEPPMDLVESQRTTVDIVRIADNMRPELAASKLTESLRLYDSYHAELFW 185
D 121 PFDSPMTWHLFEPLRDAVDKDVAMLDVSDQRTRLVVASQLSEVLRDDPYHAELFW 180
QY 186 WTGAFTSEGIPIHSLVSAESDRVTFGDRPPVANTDRPEFGHDSKVLVYSTYNER 245
D 181 WTSPPVLAHGVPPPTLASDARLRVLDGRDPVRSYQNRRAELADDSKVLVSTPSPDR 240
QY 246 ASILRCGEMTSAVLIDATAGLATCTTLHTLTHLASRDLVAALIGPATPOALVRVGLAP 305
D 241 ADALRCGEVSTILLECTMAGMATCTTLHTLIESDSRDIVGLTRQGEFQALIRVGIAF 300
QY 306 EMEPPPATPRRPIDEVFHV 326
D 301 PLAAPVAPTPRRPLDVSQIR 321

RESULT 6
O07X74 MYCBO PRELIMINARY; PRT; 344 AA.
AC O07X74;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein Mb3150.
GN OrderedLocustNames=Mb3150;
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium tuberculosis complex.
OX NCBI_TaxID=1765;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=AF2122/97;
RC MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;
RA Garnier T., Eigemeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Dutchoy S., Gordin S., Lacroix C., Monsempo S., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrett B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis ";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248345; CAD95242.1; -; Genomic_DNA.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 344 AA; 38476 MW; 9DADPB3495F24770 CRC64;

Query Match 52.6%; Score 904; DB 2; Length 344;
Best Local Similarity 56.1%; Pred. No. 2.6e-66;
Matches 180; Conservative 40; Mismatches 97; Indels 4; Gaps 2;

QY 10 VIKSAVOLACRAPSLHNSQPMRWIAE--DH-TVALFLDKDRVLYATDHSGREALLGCGA 65
D 1 MLNKAVALACRAPSVHNSQPMRWVAESGSEHTVHLFVNHRRTVPATDHSGRQAIIISGCA 60
QY 66 VLDHFRVMAAAGTTANVERFPNPNDLHLASIDFSPADFTVTEGHRRLADAILRRDRL 125
```

```

Db      61 VLDHRTIMDAHMOANTTRFPQVPOBQLATVEGSPIDHTAAGRNQAIILOKRRDRL 120
Qy      126 PFAEPPMDLVESQRTVTADTVRIDVIADDMPELAAASKLTESRLYDSSYHAELFW 185
Db      121 PFDSPTMWHLEFPALRAVDVDVMDLVSDQRTRLVVAQLEVLRRDDPYHALELW 180
Qy      186 WTGAFTSEGIPIHSLVSAESDRYTFGRDPVAVANTDRPEFGHDSKVLVSTYDNER 245
Db      181 WTSPPVLAHGVPPDTLASDARLTVDIGRDPVRSYQRRRAELADDKSKVLVSTSPDTR 240
Qy      246 ASLRCGEMLSAVLLDATMAGLATCTLTHITELHASRLVLAALIGQAPQALVRGAIAP 305
Db      241 ADALRCGSVSTILLECTMAGMATCTLTHLLESSDSRDYVGLTRQGEPOALIRVGIA 300
Qy      306 EMEPPATPRRPIDEVFHVH 326
Db      301 PLAAVPAPTPRRPIDSVLQIR 321

```

```

RESULT 7
Q73260 MYCPA
ID 073260 MYCPA PRELIMINARY; PRT; 333 AA.
AC 073260
DT 05-JUN-2004 (TReMBLrel. 27, Created)
DT 05-JUN-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUN-2004 (TReMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=MAP1743c;
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OC NCBI_TaxID=1770;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=k10;
RA Li L., Bannantine J., Zhang Q., Amonsin A., Alt D., Kapur V.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AEO17233; AAS04060.1; -; Genomic_DNA.
KW Complete proteome.
SQ SEQUENCE 333 AA; 36766 MW; 8F3BA527D631E03 CRC64;

```

```

Query Match 49.3%; Score 847.5; DB 2; Length 333;
Best Local Similarity 52.3%; Pred. No. 1.2e-61;
Matches 170; Conservative 45; Mismatches 109; Indels 1; Gaps 1;

Qy      8 TDVTKSAVOLACRAPSLHNSQPMRWIAEDHTVALFLDKDRLVLTDSHSGREALGCAVL 67
Db      9 TEVITAAIETLACHAPSLHNSQPMRWVAGSTGVDLFVDPRTVKSDKSGREALISGAL 68
Qy      68 DHFRVAAAAGTTANVERFPNPNDPLHLASIDFSPADFVTEGHRLRADAILRRTDRLPF 127
Db      69 DHFRVAAAAGTSSNVAQFPNPNDPLHLASIDVPTFVVAARADLAILRRTNPLPF 128
Qy      128 AEPMDLVESQRTVTADTVRIDVIADDMPELAAASKLTESRLYDSSYHAELFWMT 187
Db      129 RAPGWSALPEVLADAGRDSVALDVGPDPARPELVAAARLTALRRYDDYHHELMWT 188
Qy      188 GAPTSGIPIHSLVSAESDRYTFGRDPVAVANTDRPEFGHDSKVLVSTYDNERAS 247
Db      189 SPGRKFGIPESALVSEADRDVAVNRFPVDPIDERSAGSYDAKILVISTPEDTRAD 248
Qy      248 LTRCGEMLSAVLLDATMAGLATCTLTHITELHASRDVLAALIGQP-ATPOLVVGIAPE 306
Db      249 ALKRGVLSKRLLECTAGLATCPTVHTVTELEAGRDILQHLMDPAVPOVLIVGVEPE 308
Qy      307 MEPPATPRRPIDEVFHVAKDHR 331
Db      309 GELPEPRLTPRRPLGDVLFQFRLSDR 333

```

RESULT 8

```

OSYW27 NOCPA
ID OSYW27 NOCPA PRELIMINARY; PRT; 333 AA.
AC OSYW27
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=nfa27670;
OS Nocardia farcinica.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Nocardia;
OC NCBI_TaxID=37329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IFM 10152;
RX PubMed=15466710; DOI=10.1073/pnas.0406410101;
RA Ishikawa J., Yamashita A., Mikami Y., Hoshino Y., Kurita H., Hotta K.,
RA Shiba T., Hattori M.;
RT "The complete genomic sequence of Nocardia farcinica IFM 10152."
RL Proc. Natl. Acad. Sci. U.S.A. 101:14925-14930(2004).
DR EMBL; AP006618; BAD57614.1; -; Genomic_DNA.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 333 AA; 36173 MW; 6EBFFA9E47BD5A81 CRC64;

```

```

Query Match 36.6%; Score 629; DB 2; Length 333;
Best Local Similarity 44.9%; Pred. No. 1.5e-43;
Matches 146; Conservative 43; Mismatches 130; Indels 6; Gaps 3;

Qy      1 MPDTMTVDVYKSAVOLACRAPSLHNSQPMRWIAEDHTVALFLDKDRLVLTATDSHREAL 60
Db      12 VPDQQTLLAVMR-----AARAPSLHNTQPMRWFPDGRRLRLFFDDRLDADPNKRLV 67
Qy      61 LGCGAVLDHFRVAAAAGTTANVERFPNPNDPLHLASIDFSPADFVTEGHRLRADAILRL 120
Db      68 ISCGAMLHVRTAFAPAAAGMHTDTKRLPAPDRPDLAALFRPWPDPGVRVAAEAIIDHR 127
Qy      121 RTRRLPFAEPPMDLVESQRTVTADTVRIDVIADDMPELAAASKLTESRLYDSSYH 180
Db      128 RTRRLPDEPPDLGGLGHEHTARKLADPHDVTGVLDTPARLAAASEHSTALNHVMPYQ 187
Qy      181 AELFWMTGAFTSEGIPIHSLVSAESDRYTFGRDPVAVANTDRPEFGHDSKVLVST 240
Db      188 AELRWMTGHEGSHGVPAAHALASPDAAKVPVGRPPRPAQPSARRD-QPDRGLVLST 246
Qy      241 YDNERASLTRCGEMLSAVLLDATMAGLATCTLTHITELHASRDVLAALIGQAPQALVR 300
Db      247 AGDTVSWLHAGSALSLLECTADGATCPTLTHITELATTRQLANLHAGRAGVGVVVR 306
Qy      301 VGLAPEMEPPATPRRPIDEVHV 325
Db      307 VGVAFGNEHDHP-TPRRRPDEFLTV 330

```

```

RESULT 9
OSYV8 NOCPA
ID OSYV8 NOCPA PRELIMINARY; PRT; 335 AA.
AC OSYV8
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=nfa28960;
OS Nocardia farcinica.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Nocardia;
OC NCBI_TaxID=37329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IFM 10152;
RX PubMed=15466710; DOI=10.1073/pnas.0406410101;
RA Ishikawa J., Yamashita A., Mikami Y., Hoshino Y., Kurita H., Hotta K.,
RA Shiba T., Hattori M.;
RT "The complete genomic sequence of Nocardia farcinica IFM 10152."

```



GN OrderedLocusNames=RV3131;  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;  
 OC Mycobacterium tuberculosis complex.  
 CX NCBI\_TaxID=1773;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230; DOI=10.1038/21159;  
 RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C.M.,  
 RA Harris D.E., Gordon S.V., Eigemeier K., Gao S., Barry C.E. III,  
 RA Tekala F., Badcock K., Bauman D., Brown D., Chillingworth T.,  
 RA Connor R., Davies R.M., Devlin K., Feltwell T., Gentles S., Hamlin N.,  
 RA Holtroyd S., Hornby T., Jagers K., Krogh A., McLean J., Moule S.,  
 RA Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,  
 RA Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.,  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 DR EMBL; BX642582; CAB06283.1; -; Genomic\_DNA.  
 DR PIR; D70645; D70645.  
 DR Tuberculac; RV3131; -.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 332 AA; 35978 MW; 8C28FF93130D131A CRC64;  
 Query Match 24.3%; Score 417; DB 2; Length 332;  
 Best Local Similarity 36.1%; Pred. No. 5.2e-26;  
 Matches 120; Conservative 45; Mismatches 139; Indels 28; Gaps 11;  
 QY 2 PDVTWTTDVIKSAVOLACRAPSLHNSQPMRWIAEDHTVAFLEDKDRLVYATDHSGREALT 61  
 DB 6 PD-----AETVRTVTLAVRAPSIHNTQPMRWVCPSTLELFSRPMQLRSTDPDGRRLIL 61  
 QY 62 GCGAVLDHFRVMAAAAGTTANVERFPNNDPLHLASIDFSP-----ADFTGHRRLADA 116  
 DB 62 SCGVALHHCVVALLASLGWQAKVNRFPDPKDRCHLATIGVQPLVPDQADVAL-----AAA 115  
 QY 117 ILRRTDRLEPFA--EPPDMVLVESQLRTTADTVADTVRIDVADMRBELAASKLTESRL 174  
 DB 116 IPRRTDRRAVSCMPVPGGDIALMAAARAAGVWL-R-QVSLDRMKAIVAAVD--HV 171  
 QY 175 YDSSYHAELFMWGTAFETSGIPHSSLVSAESDRVTFGRDF--PVVAN-TDRRPERGHD 231  
 DB 172 TDEYRLBELTWSGRYSVAGVPARNEPPSDPAPIP-GRLEFGRLSGQSDVLP--ADD 228  
 QY 232 RSKVLVSTYNBERASLRCGEMLSAVLLDATMAGLATCTTHTTELHASDLY-AALIG 290  
 DB 229 GAAILAETETDRLARLRAGEAASIVLLTATMAGLACCPTEPLEIAKTRDAVRAEVFG 288  
 QY 291 QPATPOLVRYGLAPMEHEPPATPRRIDEV 322  
 DB 289 AGGYPQMLRLVGWAPINADPLPTPRRELISOV 320  
 RESULT 13  
 Q7TX70\_MYCBO PRELIMINARY; PRT; 332 AA.  
 AC Q7TX70;  
 DT 01-OCT-2003 (TREMBlrel. 25, Created)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein M3155.  
 GN OrderedLocusNames=M3155;  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;  
 OC Mycobacterium tuberculosis complex.  
 CX NCBI\_TaxID=1765;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=AF2122/97;

RX MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;  
 RA Garnier T., Eigemeier K., Camus J.-C., Medina N., Mansoor H.,  
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monempe C., Simon S.,  
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.,  
 RA "The complete genome sequence of Mycobacterium bovis."  
 RT complete genome sequence of Mycobacterium bovis."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
 DR EMBL; BX248345; CAD95247.1; -; Genomic\_DNA.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 332 AA; 35978 MW; 8C28FF93130D131A CRC64;  
 Query Match 24.3%; Score 417; DB 2; Length 332;  
 Best Local Similarity 36.1%; Pred. No. 5.2e-26;  
 Matches 120; Conservative 45; Mismatches 139; Indels 28; Gaps 11;  
 QY 2 PDVTWTTDVIKSAVOLACRAPSLHNSQPMRWIAEDHTVAFLEDKDRLVYATDHSGREALT 61  
 DB 6 PD-----AETVRTVTLAVRAPSIHNTQPMRWVCPSTLELFSRPMQLRSTDPDGRRLIL 61  
 QY 62 GCGAVLDHFRVMAAAAGTTANVERFPNNDPLHLASIDFSP-----ADFTGHRRLADA 116  
 DB 62 SCGVALHHCVVALLASLGWQAKVNRFPDPKDRCHLATIGVQPLVPDQADVAL-----AAA 115  
 QY 117 ILRRTDRLEPFA--EPPDMVLVESQLRTTADTVADTVRIDVADMRBELAASKLTESRL 174  
 DB 116 IPRRTDRRAVSCMPVPGGDIALMAAARAAGVWL-R-QVSLDRMKAIVAAVD--HV 171  
 QY 175 YDSSYHAELFMWGTAFETSGIPHSSLVSAESDRVTFGRDF--PVVAN-TDRRPERGHD 231  
 DB 172 TDEYRLBELTWSGRYSVAGVPARNEPPSDPAPIP-GRLEFGRLSGQSDVLP--ADD 228  
 QY 232 RSKVLVSTYNBERASLRCGEMLSAVLLDATMAGLATCTTHTTELHASDLY-AALIG 290  
 DB 229 GAAILAETETDRLARLRAGEAASIVLLTATMAGLACCPTEPLEIAKTRDAVRAEVFG 288  
 QY 291 QPATPOLVRYGLAPMEHEPPATPRRIDEV 322  
 DB 289 AGGYPQMLRLVGWAPINADPLPTPRRELISOV 320  
 RESULT 14  
 Q7DE27\_MYCTU PRELIMINARY; PRT; 344 AA.  
 ID Q7DE27\_MYCTU  
 AC Q7DE27;  
 DT 05-JUN-2004 (TREMBlrel. 27, Created)  
 DT 05-JUN-2004 (TREMBlrel. 27, Last sequence update)  
 DT 05-JUN-2004 (TREMBlrel. 27, Last annotation update)  
 DE Hypothetical protein.  
 GN OrderedLocusNames=MT3217;  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;  
 OC Mycobacterium tuberculosis complex.  
 CX NCBI\_TaxID=1773;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RX MEDLINE=22206494; PubMed=12218036;  
 RX DOI=10.1128/JB.184.19.5479-5490.2002;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J.D., Deboy R.T., Dodson R.J., Gail M.L., Hatt D.H.,  
 RA Hickey E.K., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D.,  
 RA Salzberg S.L., Delcher A., Uetzerback T.R., Weidman J.F., Khouri H.M.,  
 RA Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,  
 RA Fraser C.M.,  
 RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL J. Bacteriol. 184:5479-5490(2002).  
 DR EMBL; AE000516; AAK47555.1; -; Genomic\_DNA.  
 DR TIGR; MT3217; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 344 AA; 37048 MW; 6ACAD2A4C05DCB49 CRC64;



Query Match 24.3%; Score 417; DB 2; Length 344;  
 Best Local Similarity 36.1%; Pred. No. 5.5e-26;  
 Matches 120; Conservative 45; Mismatches 139; Indels 28; Gaps 11;

QY 2 PDVTWTTIVISAAGVLAARAPSLHNSQPRWTAEHHTVALFLDKRVLVYATDHSGREALL 61  
 DB 18 PD---AETVRVLTALVAPSLHNTQPRMRCVPTSLFSPRMQSRSTDPDRELIL 73  
 QY 62 GCGAVLDFRVMAMAAGTTANVERPPNPNDPLHLASIFSP-----ADFTVGHRLRADA 116  
 DB 74 SCGVALLHHCYVALALSLGQAKNRFPDPKDRCHLATTIGQPLVPGQADVAL-----AAA 127  
 QY 117 ILLRTRDRLPPA--BPPMDLVESQLRTTVDTRIVDIADDMPELAASKLTESRL 174  
 DB 128 IPRRTDRRAVSCWVPGGDILMAARAAGGVMLR-QVSLDRMKALVAQVLD--HV 183  
 QY 175 YDSSTHALFWMWGAFETSEGIPIHSLVSAESDVTGROP--PVVAN-IDRRPEFGHD 231  
 DB 184 TDEEYLRRLTWGSGRYGVAGVPARNBPPSPAPIP-GRLEAGGLSQSPDVLRF--ADD 240  
 QY 232 RSKVLVSTYDNERSALRCGEMLSAVLLDATMAGLATCTLTHITELHASRDV-ALIG 290  
 DB 241 GAIALTGETDRLRLARLRAGEASIVLLTRAMGLACCPTEPLEIAKTRDAVAEAVFG 300  
 QY 291 QPATPOLVRVGLAPEMEEPATRRPIDEV 322  
 DB 301 AGYPOMLRIVGMAPINADPLPPTRRRLSQV 332

RESULT 15  
 Q63HY5 BURPS PRELIMINARY; PRT; 343 AA.  
 AC 063HY5  
 DT 25-OCT-2004 (TRENBLrel. 28, Created)  
 DT 25-OCT-2004 (TRENBLrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TRENBLrel. 28, Last annotation update)  
 DE Hypothetical protein.  
 GN OrderedLocustNames=BPSS2284;  
 OS Burkholderia pseudomallei (Pseudomonas pseudomallei).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 CC Burkholderiaceae; Burkholderia; pseudomallei group.  
 NC NCB1 TaxID=28450;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=K96243;  
 RX PubMed=15377794; DOI=10.1073/pnas.0403302101;  
 RA Holden M.T.G., Tiltball R.W., Peacock S.J., Cerdano-Tarraga A.-M.,  
 RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.L.,  
 RA Bentley S.D., Sebaihia M., Thomson N.R., Bason N., Beacham I.R.,  
 RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,  
 RA Chillingworth T., Cronin A., Crosssett B., Davis P., Deshazer D.,  
 RA Feltham T., Frazer A., Hance Z., Hauser H., Holtroyd S., Jagels K.,  
 RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,  
 RA Rabinowitsch E., Rutherford K., Sanders M., Simmonds M.,  
 RA Songstvetill S., Stevens K., Tumapa S., Vesaratchaveest M.,  
 RA Whitehead S., Yeats C., Barrett B.G., Oyston P.C.F., Parkhill J.,  
 RT "Genomic plasticity of the causative agent of melioidosis.  
 RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).  
 DR EMBL; BX571966; CAH39770.1; -; Genomic DNA.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 343 AA; 37037 MW; D2391D16284362B4 CRC64;

Query Match 19.5%; Score 334.5; DB 2; Length 343;  
 Best Local Similarity 31.9%; Pred. No. 3.7e-19;  
 Matches 105; Conservative 49; Mismatches 148; Indels 27; Gaps 8;

QY 9 DVIKAVOLACAPSLHNSQPRWTAEHHTVALFLDKRVLVYATDHSGREALLGCGAVLD 68  
 DB 25 DKRLAIRYVALIAPSSHTQPRMRCVPTSLFSPRMQSRSTDPDRELILSCAALL 84  
 QY 69 HFRVMAAAGTTANVERPPNPNDPLHLASIFSP-----ADFTVGHRLRADA 128

DB 85 NLRVALSRLFGIAYIDMFPSPPDDVIALVRLDPHGTHDSLVLPFDAILERTVTRPIYA 144  
 QY 129 E---PPMDLVESQLRTTVDTRIVDI-ADDMPELAASKLTESRLTYDSSYHAELEF 184  
 DB 145 NEAVPCD---VQRALVDAGAAEGAEIACVDAPHALDEIADADRLQFADPPFRRELA 201  
 QY 185 WWTGAFTSEGIPIHSLVSAESDVTGROPFPVANTDRRPERG-----H---DRSK 234  
 DB 202 NWVHPRRHDDQMPAFAAGVPALDFAF-----PVVASAIRTFDGGGILAMHRLVDGSP 256  
 QY 235 VLV-LSITYDNERSALRCGEMLSAVLLDATMAGLATCTLTHITELHASRDVLAALIGQPA 293  
 DB 257 LIVGISASDDRDAMVAAQGLERVLVAAAAGLTASYLNQPIEIDALREKRLPLHLYDA 316  
 QY 294 TPQALVRVGLAPEMEEPATRRPIDEV 322  
 DB 317 HPQLLRIGRGFPVVAH---APRRPLNDV 341

Search completed: March 23, 2006, 05:24:24  
 Job time : 86.6574 secs



**THIS PAGE BLANK (USPTO)**

**This Page Blank (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OW protein - protein search, using SW model

Run on: March 23, 2006, 05:24:53 ; Search time 22.5753 Seconds  
(without alignments) \ 1212.194 Million cell updates/sec

Title: US-10-617-038-25

Perfect score: 1718  
Sequence: 1 MPDMMVTVDVKSAYOLACR.....PATRRRPIDEVFRAXDHR 331

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:  
1: /cgn2\_6/ptodata/1/1aa/5\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/6\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/H\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/PCITUS\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/RE\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfilltest.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	529	30.8	104	2	US-09-072-596-244
2	529	30.8	104	2	US-09-072-967-249
3	529	30.8	104	2	US-10-193-002-244
4	529	30.8	104	2	US-10-084-843-249
5	106.5	6.2	413	2	US-09-266-965-139
6	100.5	5.8	1176	2	US-09-976-594-793
7	96.5	5.6	415	2	US-10-758-759-23
8	95	5.5	534	2	US-10-104-047-3403
9	92.5	5.4	519	2	US-09-252-991A-16969
10	90.5	5.3	664	2	US-09-252-991A-16969
11	90.5	5.3	1595	2	US-08-978-277A-4
12	90	5.2	285	2	US-09-248-766A-20009
13	89.5	5.2	371	2	US-09-902-540-15386
14	89.5	5.2	502	2	US-09-489-039A-14063
15	89	5.2	1580	1	US-08-804-227C-11
16	89	5.2	1580	1	US-08-804-198-5
17	88.5	5.2	201	2	US-09-583-110-5297
18	88.5	5.2	206	2	US-09-107-433-2696
19	88.5	5.2	1016	2	US-09-107-533A-9330
20	88.5	5.2	1016	2	US-09-252-991A-25429
21	88.5	5.2	1346	1	US-08-635-121-2
22	88.5	5.2	1346	2	US-08-978-277A-2
23	87.5	5.1	732	2	US-09-252-991A-30074
24	87.5	5.1	979	2	US-09-323-872A-55
25	87.5	5.1	979	2	US-09-072-433-9
26	87	5.1	638	2	US-09-477-962-105
27	86.5	5.0	497	1	US-08-075-193-4

28	86.5	5.0	497	1	US-08-564-090A-4	Sequence 4, Appli
29	86.5	5.0	497	4	PCT-US94-06698-4	Sequence 4, Appli
30	86.5	5.0	522	2	US-09-949-016-9740	Sequence 9740, Ap
31	86.5	5.0	522	2	US-09-949-016-9741	Sequence 9741, Ap
32	86.5	5.0	3545	2	US-09-679-279-13	Sequence 13, Appl
33	86	5.0	484	2	US-09-605-703B-2520	Sequence 2520, Ap
34	85.5	5.0	646	2	US-09-328-352-6017	Sequence 6017, Ap
35	85.5	5.0	942	2	US-09-252-991A-30681	Sequence 30681, A
36	85.5	5.0	3413	2	US-10-042-665A-8	Sequence 8, Appli
37	85	4.9	3782	2	US-09-105-537-4	Sequence 4, Appli
38	84.5	4.9	869	2	US-09-489-039A-11429	Sequence 11429, A
39	83.5	4.9	479	2	US-09-328-352-5396	Sequence 5396, Ap
40	83	4.8	297	1	US-08-738-944-51	Sequence 51, Appl
41	83	4.8	297	2	US-09-263-352-41	Sequence 41, Appl
42	83	4.8	691	2	US-09-758-759-159	Sequence 159, App
43	83	4.8	737	2	US-09-602-787A-182	Sequence 182, App
44	82.5	4.8	370	2	US-08-369-822C-2	Sequence 2, Appli
45	82.5	4.8	370	2	US-08-582-776C-2	Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-09-072-596-244  
; Sequence 244, Application US/09072596  
; Patent No. 6458366  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yaelir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonia  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedrick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Hendrickson, Ronald C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; NUMBER OF SEQUENCES: 350  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/072,596  
; FILING DATE: 05-MAY-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Makl, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 244:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 104 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-072-596-244  
; Query Match 30.8%; Score 529; DB 2; Length 104;  
; Best Local Similarity 100.0%; Pred. No. 86-51;

Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 MAAAGTTANVERFPNPNDPLHLASIDFSPADFTVTEGHRRLRADAILRRTDRLPFAEPPDW 133  
Db 1 MAAAGTTANVERFPNPNDPLHLASIDFSPADFTVTEGHRRLRADAILRRTDRLPFAEPPDW 60

Qy 134 DLVESQRTTVTADTVRIDVADMRPELAASKLTESLRLYDS 177  
Db 61 DLVESQRTTVTADTVRIDVADMRPELAASKLTESLRLYDS 104

RESULT 2  
US-09-072-967-249

; Sequence 249, Application US/09072967  
; Patent No. 6592877

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Houghton, Raymond

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Twardzik, Daniel R.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Hendrickson, Ronald C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

; NUMBER OF SEQUENCES: 355

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/072,967

; FILING DATE: 05-MAY-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 249:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 104 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-09-072-967-249

Query Match 30.8%; Score 529; DB 2; Length 104;

Best Local Similarity 100.0%; Pred. No. 8e-51;

Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 MAAAGTTANVERFPNPNDPLHLASIDFSPADFTVTEGHRRLRADAILRRTDRLPFAEPPDW 133  
Db 1 MAAAGTTANVERFPNPNDPLHLASIDFSPADFTVTEGHRRLRADAILRRTDRLPFAEPPDW 60

Qy 134 DLVESQRTTVTADTVRIDVADMRPELAASKLTESLRLYDS 177  
Db 61 DLVESQRTTVTADTVRIDVADMRPELAASKLTESLRLYDS 104

RESULT 3  
US-10-193-002-244

; Sequence 244, Application US/10193002  
; Patent No. 6949246

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Houghton, Raymond

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Twardzik, Daniel R.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Hendrickson, Ronald C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

; TUBERCULOSIS

; NUMBER OF SEQUENCES: 350

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/193,002

; FILING DATE: 10-Jul-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/072,596

; FILING DATE: 05-MAY-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 682-4900

; INFORMATION FOR SEQ ID NO: 244:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 104 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 244:

US-10-193-002-244

Query Match 30.8%; Score 529; DB 2; Length 104;

Best Local Similarity 100.0%; Pred. No. 8e-51;

Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 MAAAGTTANVERFPNPNDPLHLASIDFSPADFTVTEGHRRLRADAILRRTDRLPFAEPPDW 133  
Db 1 MAAAGTTANVERFPNPNDPLHLASIDFSPADFTVTEGHRRLRADAILRRTDRLPFAEPPDW 60

Qy 134 DLVESQRTTVTADTVRIDVADMRPELAASKLTESLRLYDS 177  
Db 61 DLVESQRTTVTADTVRIDVADMRPELAASKLTESLRLYDS 104

RESULT 4  
US-10-084-843-249

; Sequence 249, Application US/10084843  
; Patent No. 6962710

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; Dillon, Davin C.  
 ; Campos-Neto, Antonio  
 ; Houghton, Raymond  
 ; Vedrick, Thomas S.  
 ; Twardzik, Daniel R.  
 ; Lodes, Michael J.  
 ; Hendrickson, Ronald C.  
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
 ; AND DIAGNOSIS OF TUBERCULOSIS  
 ; NUMBER OF SEQUENCES: 355  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SEED and BERRY LLP  
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue  
 ; CITY: Seattle  
 ; STATE: Washington  
 ; COUNTRY: USA  
 ; ZIP: 98104-7092  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/084,843  
 ; FILING DATE: 25-Feb-2002  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/072,967  
 ; FILING DATE: 05-MAY-1998  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Makl, David J.  
 ; REGISTRATION NUMBER: 31,392  
 ; REFERENCE/DOCKET NUMBER: 210121.411C9  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (206) 622-4900  
 ; TELEFAX: (206) 682-6031  
 ; INFORMATION FOR SEQ ID NO: 249:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 104 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 249:  
 ; US-10-084-843-249  
 ;  
 ; Query Match 30.8%; Score 529; DB 2; Length 104;  
 ; Best Local Similarity 100.0%; Pred. NO. 8e-51;  
 ; Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 ;  
 ; QY 74 MAAAGTANVERFPNPNDPLHSLADSPADPVTGHRRLADAILLRTDRLPFAEPDWM 133  
 ; DB 1 MAAAGTANVERFPNPNDPLHSLADSPADPVTGHRRLADAILLRTDRLPFAEPDWM 60  
 ; QY 134 DLVESQLRTVTADTVRIDVADMRPELAASKLTESLRLYDS 177  
 ; DB 61 DLVESQLRTVTADTVRIDVADMRPELAASKLTESLRLYDS 104  
 ;  
 ; RESULT 5  
 ; US-09-266-965-139  
 ; Sequence 139, Application US/09266965  
 ; Patent No. 6495348  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D  
 ; APPLICANT: Mao, Y  
 ; APPLICANT: Varoglu, M  
 ; APPLICANT: He, M  
 ; APPLICANT: Sheldon, P  
 ; TITLE OF INVENTION: Matchycin biosynthetic gene cluster  
 ; FILE REFERENCE: 600.456US1  
 ; CURRENT APPLICATION NUMBER: US/09/266,965  
 ; CURRENT FILING DATE: 1999-03-12

; EARLIER APPLICATION NUMBER: US 08/624,447  
 ; EARLIER FILING DATE: 1996-08-19  
 ; EARLIER APPLICATION NUMBER: PCT/US94/11279  
 ; EARLIER FILING DATE: 1994-10-06  
 ; EARLIER APPLICATION NUMBER: US 08/133,963  
 ; EARLIER FILING DATE: 1993-10-07  
 ; NUMBER OF SEQ ID NOS: 145  
 ; SOFTWARE: FASTSEQ for Windows Version 3.0  
 ; SEQ ID NO 139  
 ; LENGTH: 413  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces lavendulae  
 ; US-09-266-965-139  
 ;  
 ; Query Match 6.2%; Score 106.5; DB 2; Length 413;  
 ; Best Local Similarity 24.1%; Pred. No. 0.0073;  
 ; Matches 73; Conservative 32; Mismatches 111; Indels 87; Gaps 14;  
 ;  
 ; QY 2 PDITWTTDVIKSAVQLACRAPSLHNSQPMRWIAEDHTVALFLDKR-----VIYAT 52  
 ; DB 107 PDPRHVADPRLAGVIAQRYPDIRHPLVW-----NELKGFYDEDRRRMDYEGYRLXNL 160  
 ; QY 53 DH-----SGREALGCGAVLDHFRVMAAAG-----TTANVERFPNPNDP 92  
 ; DB 161 VHAELKRNPRNVLGGYAVVDHDPPEADADRSRLRGFWGLDQRSADVIRYWN--- 216  
 ; QY 93 LHLASIDFSPAD---FTVEGHRRLADAILLRTDRLPFAEPDWMDLVESQLRTVTADTV 149  
 ; DB 217 AHRAGADFVYVDGSSYIREGHRALPDE--FAATLK--FAVYTRK--VRSYTGLEPYWMAEW 270  
 ; QY 150 RIDVIADMRP-----ELAASKLTESLRLYDSYHAELFW-----WTG 188  
 ; DB 271 YVEPPADDRPGRGDNGEGHRTAVQNTAMWRLAESGASAFYVNPQRTGKACPGCLMRS 330  
 ; QY 189 AFETSG--IPHSLSVAASDRYTFGRDP-----VANTDRRPFGRHRSKLYV 237  
 ; DB 331 THLRDGGGELPMAGLLS-----RFARERPPGTAFRPVAVTCGSGDVEALADEAALV 383  
 ; QY 238 LST 240  
 ; DB 384 VNT 386  
 ;  
 ; RESULT 6  
 ; US-09-976-594-793  
 ; Sequence 793, Application US/09976594  
 ; Patent No. 6673549  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Furness, Michael  
 ; APPLICANT: Buchbinder, Jenny  
 ; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
 ; FILE REFERENCE: PA-0041 US  
 ; CURRENT APPLICATION NUMBER: US/09/976,594  
 ; PRIOR FILING DATE: 2001-10-12  
 ; PRIOR APPLICATION NUMBER: 60/240,409  
 ; NUMBER OF SEQ ID NOS: 1143  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 793  
 ; LENGTH: 1176  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: Incyte ID No. 6673549 1806212CD1  
 ; US-09-976-594-793  
 ;  
 ; Query Match 5.8%; Score 100.5; DB 2; Length 1176;  
 ; Best Local Similarity 22.0%; Pred. No. 0.19;  
 ; Matches 57; Conservative 38; Mismatches 87; Indels 77; Gaps 13;  
 ;  
 ; QY 86 FPND--NDPLHLA--SIDFSPADPVTGHRRLADAIL-----LRRTR 124  
 ; DB 1176

Db 50 PYPYPMNGRIHLGHFFSLSKCEFAVGYQYLKKGKCLFPFGHHTQMPYACADKLRETE 109

Qy 125 LPFAEPPDM-DIVESQRLTTVTAQIVRIDVADNMEPELAA---ASK---LTESRLDYD 176

Db 110 L-YGCPDPDPDEEEEEEESTSVKTEDIIIKOKAKGKSKAAAKAGSSKYQWIKMSLGLSD 168

Qy 177 S---SYHAELEFV-----WTGAFFETSEGIPIHS-----SLVSAAS 207

Db 169 EIVYESSAEHMLDYFPPLAIDDLKRMGLKVDPRRSFTTIDVNPYDSFVRWQFLTLRER 228

Qy 208 DRVTFGRDFPVVANYNDRRPEFGHDS-----KVLVLSTYDNERASILRCGR- 253

Db 229 NKIKGKKRYTITSPPDGGRCMDHNRQITGEGVGRQEVYTLKLAKLBYPPSLSLK--GKN 286

Qy 254 --MISAVLIDATMAGLATC 270

Db 287 IFLVAAATLRPETMFGQTC 305

RESULT 7  
 US-09-758-759-23  
 Sequence 23, Application US/09758759  
 Patent No. 6861513  
 GENERAL INFORMATION:  
 APPLICANT: Hosted, Thomas J.  
 APPLICANT: Wang, Tim X.  
 APPLICANT: Horan, Ann C.  
 TITLE OF INVENTION: Evernomycin Biosynthetic Genes  
 FILE REFERENCE: ID0981K US  
 CURRENT APPLICATION NUMBER: US/09/758,759  
 CURRENT FILING DATE: 2001-01-11  
 PRIOR APPLICATION NUMBER: US 60/175,751  
 PRIOR FILING DATE: 2000-01-12  
 NUMBER OF SEQ ID NOS: 204  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 23  
 LENGTH: 415  
 TYPE: PRN  
 ORGANISM: Micromonospora carbonacea  
 FEATURE:  
 OTHER INFORMATION: evdk  
 US-09-758-759-23

Query Match	5.6%	Score 96.5	DB 2	Length 415
Beet Local Similarity	22.8%	Pred. No. 0.095		
Matches	74	Conservative 29	Mismatches 140	Indels 81
				Gaps 15
Qy	17	LACR-APSINHSOPWRMTAEDHTVALFDKORVLYATHSGREALGGCAVLDPFRVAMA	75	
Db	51	LAALSLPALAQQGQERVRERE---LLIDVLSALAGVDRPFVYLASSGAV-----	99	
Qy	76	AAGTTANVEREPNPDPDLHLASIDFSPADPVTGHRLLADAILLRITRL-----	125	
Db	100	-----YPTVWPPYHERSAT- GPASAYGRA- KLRLEOELLNHTDRVQPVVTRLNV	148	
Qy	126	--PRAEP-PRMDVLYESQRLTVYADTVI----DYIADMPRELAAASKLTESLRLYSS	178	
Db	149	YGPQRPLPPGGVSHWLEATVTRGPIRLFDQPAVRVYVYVDVDTALMEVYIAQAGGD	208	
Qy	179	YHAELFWMTGAFETSEGIPIHSLVSAASDSVTFGRDPPVYANTDRRDEFGHDSKVLVL	238	
Db	209	RDR----LPTVVNVGSGLP-TSLLELLDTMSTVAGRELEV--RDVRQGFDR-----	254	
Qy	239	STYDNERASILRCGEMLSAVILLDTAMGLATCTTLTHITELHASRDLVALI-----	291	
Db	255	-----GNMLDTTLARELIGQQAISLP-----DGVRCQCAEVLVTRAGPGGS	296	
Qy	292	PATPOLVRVGLAPMEHEPPATP	315	
Db	297	PAPSA--RLGRASGRPPQPRP	318	

## RESULT 8

```

US-10-104-047-3403
; Sequence 3403 Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HEILIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIORITY FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIORITY FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3403
;
; LENGTH: 354
;
; TYPE: PRT
;
; ORGANISM: Homo sapiens
US-10-104-047-3403

```

Query Match	5.5%	Score 95;	DB 2;	Length 534;
Best Local Similarity	25.2%	Pred. No. 0.21;		
Matches	41;	Conservative	24;	Mismatches 52; Indels 46; Gaps 6
Oy	187	TGAFETSGCLPHSSLSVAESDRVTGCRDFPVNANTDRPEFGHDR-SKVLVLSTYDNER	245	
	:			
	:			
	:			
Db	97	SSAIVSSSGYPPSL-----GR-----AITGQAGFNDLSKVGISIEGGM	140	
Oy	246	ASLLRCGEMLSAV-----LDDATMGATCTVTHITELHASDVLVALLIGPATPOL	298	
	:			
	:			
	:			
Db	141	TGLATGGDLTAATVAKTVGTALSSGQMSIATNSVPVSAAPKPTSWAIAIRKPAKQPK	200	
Oy	299	VR-----VGLAPEMEPP-----ATRPPI	319	
	:			
	:			
Db	201	LKPGKNGVIGSAVPPPIKNNNIGTWDEKGSVVKAPPTPV	243	

```

RESULT 9
US-09-252-991A-30815
; Sequence 30815, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30815
; LENGTH: 519
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30815

```

	Query Match	5.4%; Score 92.5; DB 2;	Length 519;
	Similarity	23.9%; Pred. No. 0.38;	
Batches	75; Conservative	36; Mismatches	120; Indels 83; Gaps 15;
Oy	57 REALLGGCAVLDHFRVAMAAAGTTAN--VEREPEND--PLHL---ASIDFSPADFTVEGH	110	
	:   :   :   :   :   :   :   :		
Dd	1 RQTEGQGVDNAEANGIPETAAGTHQRVYARRDEBADDHPGLFRAGQGVDDLLANDLDRPGH	60	
Oy	111 RL-RADAILLRTDRLRFAPEPPMDL-----VESQLRTVTADTVRIIVITADMREPL-	162	
	: : : : : : : : : : : : : : : : :		
Dd	61 QVALDDBALARRTGCPQAVFPPTQVGCGFGEISFLACIALARRARQRRLDIGAQQSGSEVL	120	
Oy	163 -----AAASKLTSLKLYDSSTHAELFWTGTGAETSGIRPHSSLSVAAL	206	
	: : : : : : : : : : : : : : : :		
Dd	121 GTQCABERTTHPENGSAAGDALAGSQAQLHQYGAEVRVOLDAEDLND-----VALE	173	





```

CORRESPONDENCE ADDRESS:
ADDRESS: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1580 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-804-227C-11

Query Match      5.2%; Score 89; DB 1; Length 1580;
Best Local Similarity 22.8%; Pred. No. 5.7;
Matches 88; Conservative 44; Mismatches 120; Indels 134; Gaps 23;

QY 28 QPWNIAEDHYALFLDQDRVLYATDHSGRALGCGAVLDHFRVMAAAGTTANVERFP 87
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 709 EPW---AERMSVA-----AVNGPASYVAGDARALBEFGGRLSAG---VLRWP 751
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 88 NPNDPLHASTIDSPADFTVEGH-----RLRADAI---LIRRT--DRLPF-----AEP 130
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 752 -----LAGVDFA-----GHSPOVEQFRAELDLDTLGTVPATARLPFESTVTAALHE 797
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 131 PD-----W-----DLVE--SOLRTTTPAD-----TVRIDVIADDMRPELAA 164
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 798 PEGLDAAIYWRMRPERVERASTLRLLRSGHRTFVEMGPHPLGAAIDEVAAEGVHATA 857
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 165 ASKL--TESLRLYDSY----HAELFWMTGAFETSEG-----IPHS- 199
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 858 LATLRGSGGLDRFRSSVGAAPAHGVVDMDALFEGSGARVPLPTVAFSRDRVWLPTAI 917
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 200 --SLVSAESD-----RVTEGRDPFVYANTDRRPEFGHDSKVLVLST-----YDN 243
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 918 GRRAYEAAPVDAASAPGRYRVTW--TPVASDSGRPS-----GRWLIVQTPTGTADEADT 969
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 244 BRASLLRCGEMLSAVLDTATWAGLATCTLTHITELHASRDVLAALIGOPATPOLVRVGL 303
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 970 AASALGAAGVVERCLDPTTEAARVTLT-ERLAELDAQPEGLAGVVLBGRPGSTAPADA 1028
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 304 AP-----EMEEPATPRRPPI 319
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 1029 SPLDPGTAALLVQAVPDPAAPKARI 1054
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

Search completed: March 23, 2006, 05:30:15  
Job time : 23.6753 secs



This Page Blank (uspto)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 06:14:33 ; Search time 69.6072 Seconds  
(without alignments)  
1986.885 Million cell updates/sec

Title: US-10-617-038-25

Perfect score: 1718  
Sequence: 1 MPDMMVTVDVKSANVQLACR.....PATRRPRIDEVHVRADHR 331

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_Main:

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1718	100.0	331	US-10-617-038-25	Sequence 25, Appl
2	915	53.3	344	US-10-617-038-36	Sequence 36, Appl
3	529	30.8	104	US-10-193-002-244	Sequence 244, App
4	529	30.8	104	US-10-084-843-249	Sequence 249, App
5	529	30.8	104	US-11-028-898-249	Sequence 249, App
6	529	30.8	104	US-11-082-005-244	Sequence 244, App
7	417	24.3	332	US-10-617-038-40	Sequence 40, Appl
8	106.5	6.2	413	US-09-953-348-139	Sequence 139, App
9	106.5	6.2	413	US-10-267-255-139	Sequence 139, App
10	104	6.1	380	US-10-282-122A-48196	Sequence 48196, A
11	102	5.9	441	US-10-156-761-12910	Sequence 12910, A
12	99.5	5.7	441	US-10-437-965-177042	Sequence 177042, A
13	98	5.8	3352	US-10-156-761-7961	Sequence 7961, Ap
14	97.5	5.7	1094	US-10-156-761-11435	Sequence 11435, A
15	97	5.6	1354	US-10-156-761-8425	Sequence 8425, Ap
16	96.5	5.6	415	US-09-758-759-23	Sequence 23, Appl
17	96.5	5.6	415	US-11-021-825-23	Sequence 23, Appl
18	96.5	5.6	423	US-10-107-431-31	Sequence 31, Appl
19	95	5.5	534	US-10-104-004-3403	Sequence 3403, Ap
20	95	5.5	543	US-10-094-749-2430	Sequence 2430, Ap
21	94.5	5.5	319	US-10-437-963-202037	Sequence 202037, A
22	94.5	5.5	813	US-10-156-761-10119	Sequence 10119, A
23	94.5	5.5	6238	US-10-343-720-71	Sequence 71, Appl
24	94	5.4	378	US-10-282-122A-67584	Sequence 67584, A
25	93.5	5.4	198	US-09-769-744A-30	Sequence 30, Appl
26	93.5	5.4	201	US-10-472-928-1134	Sequence 1134, Ap
27	93.5	5.4	745	US-10-732-923-19911	Sequence 19911, A

28	93.5	5.4	915	4	US-10-156-761-10050	Sequence 10050, A
29	93	5.4	431	4	US-10-087-192-1644	Sequence 1644, Ap
30	93	5.4	761	4	US-10-156-761-7913	Sequence 7913, Ap
31	93	5.4	932	4	US-10-282-122A-62242	Sequence 62242, A
32	92.5	5.4	218	4	US-10-156-761-10291	Sequence 10291, A
33	92.5	5.4	1360	5	US-10-732-923-70591	Sequence 20591, A
34	92.5	5.4	1444	5	US-10-732-923-17028	Sequence 17028, A
35	92.5	5.4	7349	4	US-10-314-657-46	Sequence 46, Appl
36	92.5	5.4	7349	5	US-10-473-193-46	Sequence 46, Appl
37	91.5	5.3	750	4	US-10-282-122A-62636	Sequence 62636, A
38	91.5	5.3	750	4	US-10-282-122A-64569	Sequence 64569, A
39	91.5	5.3	1016	4	US-10-156-761-10696	Sequence 10696, A
40	90.5	5.3	421	5	US-10-732-923-11413	Sequence 11413, A
41	90.5	5.3	457	4	US-10-282-122A-63748	Sequence 63748, A
42	90.5	5.3	841	4	US-10-282-122A-54009	Sequence 54009, A
43	90.5	5.3	1596	3	US-09-902-432-4	Sequence 4, Appl1
44	90	5.2	321	4	US-10-451-467A-428	Sequence 428, App
45	90	5.2	525	6	US-11-097-143-22077	Sequence 22077, A

#### ALIGNMENTS

RESULT 1  
US-10-617-038-25  
; Sequence 25, Application US/10617038  
; Publication No. US20040057963A1  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Peter  
; APPLICANT: Rosenkrands, Ida  
; APPLICANT: Strlym, Anette  
; TITLE OF INVENTION: Therapeutic TB Vaccine  
; FILE REFERENCE: SSI5AUSA  
; CURRENT APPLICATION NUMBER: US/10/617, 038  
; CURRENT FILING DATE: 2003-07-11  
; PRIOR APPLICATION NUMBER: DK PA 2002 01098  
; PRIOR FILING DATE: 2002-07-13  
; PRIOR APPLICATION NUMBER: US 60/401, 725  
; PRIOR FILING DATE: 2002-08-07  
; NUMBER OF SEQ ID NOS: 187  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 25  
; LENGTH: 331  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-617-038-25

QY	1	MPDMMVTVDVKSANVQLACRPSLHNSQPMRWIEDHTVLFUKDKRVLYATDHSGREAL	60
DB	1	MPDMMVTVDVKSANVQLACRPSLHNSQPMRWIEDHTVLFUKDKRVLYATDHSGREAL	60
QY	61	LGCAVADHPFVAAAAGTTANVERFPNPDPPLHLASIDFSPADFTGHRRLADATILR	120
DB	61	LGCAVADHPFVAAAAGTTANVERFPNPDPPLHLASIDFSPADFTGHRRLADATILR	120
QY	121	RTDRLPFAEPDPMDLVESQRTTATVADTVRIDVIADMDRPELAASKLTESRLYDSSYH	180
DB	121	RTDRLPFAEPDPMDLVESQRTTATVADTVRIDVIADMDRPELAASKLTESRLYDSSYH	180
QY	181	AELPMWVGAFSTSGISPHSSIVSAESDRYFGDFPVVANTDRRPERGHRSLVLT	240
DB	181	AELPMWVGAFSTSGISPHSSIVSAESDRYFGDFPVVANTDRRPERGHRSLVLT	240
QY	241	YDNERASILRCGEMLSAVLADATVAGATCTLTITELHLSRDLVAALIGPATPOLV	
DB	241	YDNERASILRCGEMLSAVLADATVAGATCTLTITELHLSRDLVAALIGPATPOLV	
QY	301	VGLAPEWEEPPATRRPRIDEVHVRADHR 331	
DB	301	VGLAPEWEEPPATRRPRIDEVHVRADHR 331	

Db 301 VGLAPEMEBPPATPRRPIDEVFHYRAKDHR 331

## RESULT 2

US-10-617-038-36

; Sequence 36, Application US/10617038  
; Publication No. US20040057963A1

## GENERAL INFORMATION:

; APPLICANT: Andersen, Peter  
; APPLICANT: Rosenkrands, Ida  
; APPLICANT: Stryhn, Anette  
; TITLE OF INVENTION: Therapeutic TB Vaccine  
; FILE REFERENCE: SS15AUSA  
; CURRENT APPLICATION NUMBER: US/10/617,038  
; CURRENT FILING DATE: 2003-07-11  
; PRIOR APPLICATION NUMBER: DK PA 2002 01098  
; PRIOR FILING DATE: 2002-07-13  
; PRIOR APPLICATION NUMBER: US 60/401,725  
; PRIOR FILING DATE: 2002-08-07  
; NUMBER OF SEQ ID NOS: 187  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 36  
; LENGTH: 344  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-617-038-36

Query Match 53.3%; Score 915; DB 4; Length 344;

Best Local Similarity 56.7%; Pred. No. 3.2e-83;

Matches 182; Conservative 39; Mismatches 96; Indels 4; Gaps 2;

QY 10 VIKSAVOLACRAPELHNSQPMWIAE--DH-TVALPLDKDRVLYATDHSGERALLCGGA 65

Db 1 VLKNAVLACAPSVHNSQPMWVAESGSEHTVHLFNNRRTPATDHSROAIIISGGA 60

QY 66 VLDFRVAAMAAAGTTANVERPNNPNDPLHLASIDFSPADFTTEGHRRLADAIIILRTDRL 125

Db 61 VLDFRLAMTAAHQAQNIITRPPQNGPDQLATVEFSPIDHTTAQORRAQAIIILQRRDRL 120

QY 126 PFAEPMDLVESQRTTVAIDTVRIDVIADMRPELAASKLTESRLYDSYHAELFW 185

Db 121 PFDSPMWHLPFPALRAVDKDVAMLDVSDQRTLVVASQLSEVLRDDPYHALEW 180

QY 186 WTGAFFESGEPHSLVSAASDRTVGRDPVYANDRRPEFGHDSKTLVLTYNDR 245

Db 181 WTSFVLAHGVPPPTLASDARLRLVDLGRDPVRSYORRAELADDSKVLVLTSPDTR 240

QY 246 ASLRGCEMTSAVLDATMGATCTLTHTLHASRDVLALIGPATPOLVVGILAP 305

Db 241 ADALRCGEVSTILLECTMAGMATCTLTHTLIESDSRDIVAGLTKRGEPQALIRVGILAP 300

QY 306 EMEBPPATPRRPIDEVFHYR 336

Db 301 PLAAVPAPTPRRPIDEVFHYR 331

## RESULT 3

US-10-193-002-244

; Sequence 244, Application US/10193002  
; Publication No. US20030135026A1

## GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Hendrickson, Ronald C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; TUBERCULOSIS  
; NUMBER OF SEQUENCES: 350

## CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/193,002

FILING DATE: 10-Jul-2002

CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072,596

FILING DATE: 05-MAY-1998

## ATTORNEY/AGENT INFORMATION:

NAME: Mark, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.417C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

## INFORMATION FOR SEQ ID NO: 244:

SEQUENCE CHARACTERISTICS:

LENGTH: 104 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 244:

US-10-193-002-244

Query Match 30.8%; Score 529; DB 4; Length 104;

Best Local Similarity 100.0%; Pred. No. 4.8e-45;

Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 MAAAGTTANVERPNNPNDPLHLASIDFSPADFTTEGHRRLADAIIILRTDRLPFAEPDW 133

Db 1 MAAAGTTANVERPNNPNDPLHLASIDFSPADFTTEGHRRLADAIIILRTDRLPFAEPDW 60

QY 134 DLVESQRTTVAIDTVRIDVIADMRPELAASKLTESRLYDS 177

Db 61 DLVESQRTTVAIDTVRIDVIADMRPELAASKLTESRLYDS 104

## RESULT 4

US-10-084-843-249

; Sequence 249, Application US/10084843  
; Publication No. US20030143243A1

## GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Hendrickson, Ronald C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; AND DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 355  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092

APPLICATION NUMBER: US/09/072,596  
FILING DATE: 05-MAY-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.417C9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 244:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 104 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 244:  
US-11-082-005-244

Query Match 30.8%; Score 529; DB 6; Length 104;  
Best Local Similarity 100.0%; Pred. No. 4,8e-45;  
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 MAAAGTTANVERFPNPNPDLHLASIDFSPADFTGHRRLADAILLRTDRLPFAPEPDW 133  
Db 1 MAAAGTTANVERFPNPNPDLHLASIDFSPADFTGHRRLADAILLRTDRLPFAPEPDW 60  
Qy 134 DLVSSQLRTYTTADTVRIDVADDMRPELAASKLTESTRLYDS 177  
Db 61 DLVSSQLRTYTTADTVRIDVADDMRPELAASKLTESTRLYDS 104

## RESULT 7

US-10-617-038-40  
Sequence 40, Application US/10617038  
Publication No. US20040057963A1  
GENERAL INFORMATION:  
APPLICANT: Andersen, Peter  
APPLICANT: Rosenkrands, Ida  
APPLICANT: Stryhn, Aneette  
TITLE OF INVENTION: Therapeutic TB Vaccine  
FILE REFERENCE: SSISUSA  
CURRENT APPLICATION NUMBER: US/10/617,038  
CURRENT FILING DATE: 2003-07-11  
PRIOR APPLICATION NUMBER: DK PA 2002 01098  
PRIOR FILING DATE: 2002-07-13  
PRIOR APPLICATION NUMBER: US 60/401,725  
PRIOR FILING DATE: 2002-08-07  
NUMBER OF SEQ ID NOS: 187  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 40  
LENGTH: 332  
TYPE: PRT  
ORGANISM: Mycobacterium tuberculosis  
US-10-617-038-40

Query Match 24.3%; Score 417; DB 4; Length 332;  
Best Local Similarity 36.1%; Pred. No. 4.9e-33;  
Matches 120; Conservative 45; Mismatches 139; Indels 28; Gaps 11;

Qy 2 PDVTWTTDIVIKSAVOLACRAPSLNSQPRWIAEDHTVALFLDKDRVLATDHSGRBAL 61  
Db 6 PD-----AETVATVTLAVRAPSHINTQPMKRWCVCTSELFSRDMQLRSTDPGRRLIL 61  
Qy 62 GCGAVLDHFRVMAAAGTTANVERFPNPNPDLHLASIDFSP-----ADFTEGHRRLADA 116  
Db 62 SCGVALLHCVVALLASLGQAKVNFPPKDRCHLATIGVQPLVDPQADVAL-----AAA 115  
Qy 117 ILARTRRLPRA--EPPMDLVESQLRTYTTADTVRIDVADDMRPELAASKLTESTRL 174  
Db 116 IPRRTTRRAYSQMPVPGDITALMAARAAGCVMLR-QVSLDLMKALIVAQADVD--HV 171  
Qy 175 YDSSYHAELFWMTGAFETSEGIPIHSSILVSAESDRVTFGDF--PVVAN--TDRRPEFGHD 231

Db 172 TDEEYLAELTITMSGRYGVSVAGVPARNPEPSPAPRIP-GRFLFAGPGISQPSDVLFP--ADD 228  
Qy 232 RSKVLVSTYDNDNRASLRGEMLSAVILDATMAAGLATCTLTITELHMSRDLY-ALLIG 230  
Db 229 GAAILAAGTSTDRLAFLRAGEAASIVLLTATMGLACCPITBPLETAKTRADAVRAEVFG 288  
Qy 291 OPATPOLVRVGLAPEMEBPPATPRRPRIDEV 322  
Db 289 AGGYFQWLRLRVGWAFTIADPLPTPRRELSCV 320

## RESULT 8

US-09-953-348-139  
Sequence 139, Application US/09953348  
Publication No. US20030134398A1  
GENERAL INFORMATION:  
APPLICANT: Sherman, David. H  
APPLICANT: Mao, Yingqing  
APPLICANT: Varoglu, Mustafa  
APPLICANT: He, Min  
APPLICANT: Sheldon, Paul  
TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER  
FILE REFERENCE: 600,530US1  
CURRENT APPLICATION NUMBER: US/09/953,348  
CURRENT FILING DATE: 2001-09-12  
PRIOR APPLICATION NUMBER: PCT/US00/06394  
PRIOR FILING DATE: 2000-03-10  
PRIOR APPLICATION NUMBER: 09/266965  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 153  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 139  
LENGTH: 413  
TYPE: PRT  
ORGANISM: Streptomyces lavendulae  
US-09-953-348-139

Query Match 6.2%; Score 106.5; DB 3; Length 413;  
Best Local Similarity 24.1%; Pred. No. 0.14;  
Matches 73; Conservative 32; Mismatches 111; Indels 87; Gaps 14;

Qy 2 PDVTWTTDIVIKSAVOLACRAPSLNSQPRWIAEDHTVALFLDKDR-----VLAT 52  
Db 107 PDRRHVADFAFLAGVLAQRYPDIRHFLW-----NELKGFYEDRRMRWYEGVTRLYNL 160  
Qy 53 DH-----SGREALGCGAVLDHFRVMAAAG-----TTANVERFPNPNP 92  
Db 161 VHAELKRRNPRNIVGGPYAVVDHPAEDDAADRSRELRCGPGELDORSADVIRYWN---- 216  
Qy 93 LHLASIDFSPAD---FVTEGHRRLADAILLRTDRLPFAPEPPMDLVESQLRTYTTADTV 149  
Db 217 AHKAGADFEVVVDSSYTRBSGHRALPDE--FAATEK--FADVTRW--VRSVTGLPVMWAEM 270  
Qy 150 RIDVIADMRP-----ELAAASKLTESTRLYDSSYHAELFW-----WTG 188  
Db 271 YVERPAEDDRPGRGDMSGBEHRTAQATAMWRLAESGASAAFTWNRQRTKACRGCLMRS 330  
Qy 189 AFETSEG--IPHSSILVSAESDRVTFGDFP-----VVANTDRRPEFGHRSKVLV 237  
Db 331 THLRDGGGELPMAGLLS-----RFAREPPGTAFRRPVAVTSGSDRVBALDEAAVLV 383  
Qy 238 LST 240  
Db 384 VNT 386

## RESULT 9

US-10-267-255-139  
Sequence 139, Application US/10267255  
Publication No. US20030124689A1  
GENERAL INFORMATION:  
APPLICANT: Sherman, D

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/084,843  
FILING DATE: 25-FEB-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/072,967  
FILING DATE: 05-MAY-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.411C9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 249:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 104 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 249:  
US-10-084-843-249  
Query Match 30.8%; Score 529; DB 4; Length 104;  
Best Local Similarity 100.0%; Pred. No. 4.8e-45;  
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 74 MAAAGTTANVERFPNPNDPLHLASIDFSPADFTVEGHRLRADAILLRTRRLPFAEPPDW 133  
DB 1 MAAAGTTANVERFPNPNDPLHLASIDFSPADFTVEGHRLRADAILLRTRRLPFAEPPDW 60  
DB 134 DLVESQLRTVTYADTVRIDVIADDMRPPELAASKLTESLRLYDS 177  
DB 61 DLVESQLRTVTYADTVRIDVIADDMRPPELAASKLTESLRLYDS 104  
RESULT 5  
US-11-028-898-249  
Sequence 249, Application US/11028898  
Publication No. US20050136069A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
Skelky, Yasir A.W.  
Dillon, Davin C.  
Campos-Neto, Antonio  
Houghton, Raymond  
Vedvick, Thomas S.  
Twardzik, Daniel R.  
Lodes, Michael J.  
Hendrickson, Ronald C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
AND DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 355  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/11/028,998

FILING DATE: 03-Jan-2005  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/10/084,843  
FILING DATE: 03-Jan-2005  
APPLICATION NUMBER: US/09/072,967  
FILING DATE: 05-MAY-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.411C9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 249:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 104 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 249:  
US-11-028-898-249  
Query Match 30.8%; Score 529; DB 6; Length 104;  
Best Local Similarity 100.0%; Pred. No. 4.8e-45;  
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 74 MAAAGTTANVERFPNPNDPLHLASIDFSPADFTVEGHRLRADAILLRTRRLPFAEPPDW 133  
DB 1 MAAAGTTANVERFPNPNDPLHLASIDFSPADFTVEGHRLRADAILLRTRRLPFAEPPDW 60  
QY 134 DLVESQLRTVTYADTVRIDVIADDMRPPELAASKLTESLRLYDS 177  
DB 61 DLVESQLRTVTYADTVRIDVIADDMRPPELAASKLTESLRLYDS 104  
RESULT 6  
US-11-082-005-244  
Sequence 244, Application US/11082005  
Publication No. US20050181419A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
Skelky, Yasir A.W.  
Dillon, Davin C.  
Campos-Neto, Antonia  
Houghton, Raymond  
Vedvick, Thomas S.  
Twardzik, Daniel R.  
Lodes, Michael J.  
Hendrickson, Ronald C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
TUBERCULOSIS  
NUMBER OF SEQUENCES: 350  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/11/082,005  
FILING DATE: 15-Mar-2005  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/10/193,002  
FILING DATE: 10-Jul-2002

```
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600,456US1
; CURRENT APPLICATION NUMBER: US/10/267,255
; PRIOR FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 09/266,965
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 08/624,447
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: PCT/US94/11279
; PRIOR FILING DATE: 1994-10-06
; PRIOR APPLICATION NUMBER: US 08/133,963
; PRIOR FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 139
; LENGTH: 413
; TYPE: PRF
; ORGANISM: Streptomyces lavendulae
US-10-267-255-139
```

```
Query Match          6.2%; Score 106.5; DB 4; Length 413;
Best Local Similarity 24.1%; Pred. No. 0.14;
Matches 73; Conservative 32; Mismatches 111; Indels 87; Gaps 14;
```

```
QY 2 PGTMTTIVIKSAVOLACRAPSLHNSQPRWIAEDHTVALFLDKCR-----VLYAT 52
DB 107 PDPRIYADPARIAGIYAIRYPDIRHFLW-----NELKGFDEDRRMDYEGYTRLNYL 160
QY 53 DH-----SGREALIGCAVLDHFRVMAAAG-----TTANVERPNNPDP 92
DB 161 VHAELKRRNPRNLVGGPYAVVDHDPADPADRSREELGPMGELDORSADVTIRWN---- 216
QY 93 LHLASIDSPAD--FVTEGHLRADAILRRTDRLPFAEPDMDLVESQRTTYTADTV 149
DB 217 AHKGADPVVVDGSGSYREGHRAIPDE--PAATEK--FADYTRW--VRSVGLPVMWMEW 270
QY 150 RIDVIADMRP-----ELAAASKLTESRLVDSSTHAELFW-----WTG 188
DB 271 YVEPAEDDRPGCDGMEGHRTAVOATAMKRLAESGASAFYVMPORGTAKCPGCLWRS 330
QY 189 AFERSSEG--IPHSLSVSAASDRVTFGRDP-----VWANDRRBPFGHDSKYLAV 237
DB 331 THLDGGSELPMAGLLS-----RFAKEFPFGTAFRVAAYTCGSGDRVEALADEAAVLY 383
QY 238 LST 240
DB 384 VNT 386
```

```
RESULT 10
US-10-282-122A-48196
; Sequence 48196, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haseelbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zuekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
```

```
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48196
; LENGTH: 380
; TYPE: PRF
; ORGANISM: Burkholderia cepacia
US-10-282-122A-48196
```

```
Query Match          6.1%; Score 104; DB 4; Length 380;
Best Local Similarity 22.2%; Pred. No. 0.21;
Matches 72; Conservative 41; Mismatches 118; Indels 94; Gaps 15;
```

```
QY 11 IKSAVOLACRAPSL-----HNSQPRWIAEDHTVALFLDKDRVYATDHSREALIGCGA 65
DB 95 VTTGISADPRAHTATATVAHDVREHIVQGHVPPILAQPGCVLRAGHT--EA--GCDP 150
QY 66 VLDFRVMAAAGTANVERPNNPNDPLHLASIDFSPADFVTEGHLRADAILRRTDRL 125
DB 151 T-----ALAGLT-----PAAYCE--IKDDGTMARLDLI 179
QY 126 PFAEPDMDLVESQRTTYTADTVRIDVIADMRBELAASKLTESRLVDSSTHAELFW 185
DB 180 EFAR-----EHLKIGTIADLIQYRSRTESIIRIA-----ERTWQTAHGTRAVLY- 226
QY 186 WTGAFTSEGIIPHSLSVSAASDRVTFGRDPFPVAVANTDRRPFEGHDSKYLVLSTYDNER 245
DB 227 ---RDQPSGSPHIALVRCAPSPDV---DTPVAV-----HEPLSVL----- 260
QY 246 ASILRCGEMLSAVLLDATMGATCTLTHITELHASRD--VAALIGPAPFOALVRVGLA 304
DB 261 -DLLETGSTHSTWTLDAARDI-----AERDLGVILNCGDTKEHLIDVFKA 307
QY 305 PEMEPPPATPRRPID-EVFHVRK 328
DB 308 FDEBEKAAALKRRPVDKFTGIGNQ 332
```

```
RESULT 11
US-10-156-761-12910
; Sequence 12910, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, WASHIIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
```



This Page Blank (uspto)



Db 2373 -----FLRSVSELSHGAVVDLTAPRAGRPV 2398  
QY 294 TPQALVRVGLAP-----EMEEPPPA 313  
Db 2399 RRVDLPTTAFQFQRYWLESADPVPFA 2423

```

RESULT 14
US-10-156-761-11435
Sequence 11435, Application US/10156761
Publication NO. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIDA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, KASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 11435
LENGTH: 1094
TYPE: PR1
ORGANISM: Streptomyces avermitilis
US-10-156-761-11435

```

Query Match	5.7%	Score 97.5;	DB 4;	Length 1094;
Beet Local Similarity	23.5%	Pred. No. 4.4;		
Matches	51;	Mismatches	64;	Indels 67; Gaps 13;
Conservative	35;			
Qy	126	PFAPBPDDLV-----ESQRTTVAITVRI--DVIAD-----DMREP-----AAA	165	
Db	460	PAEDPDWDVAAPVALPRDAVEVGAGADTVRYLAIEVLPSLLPAGLRERRVELRTIGVAR	519	
Qy	166	SKTRESL-RLYDSSSTHAELFWTTGAFERSEGI PHSLSVSAASBPDTVTCGRGPPVYANDR	224	
Db	520	VPLTEAVRL--AGLEKDPGMMRLRLYDLAG-----DPEELSL-GLPVL--ADG	564	
Qy	225	RPEFGHRSKYLIVLSTYDNERA---SLRCSEMTSAVLIDATMGATCTCTHTHTELHAS	281	
Db	565	RTTIG--PRQVILPTADGPQAPAEALRLG-----LKVHNPEAA	601	
Qy	282	RDVLAALLGQATPQAL-----YKRGAPMBEERPPA	313	
Db	602	HPLEKLGALPATERPAVLTTPQVRAAVALASLDDENAPA	638	

RESULT 15  
US/10-156-761-8425  
Sequence 8425, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHITAKA  
APPLICANT: HATTORI, MASASHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156, 761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089

```

: PRIOR FILING DATE: 2001-05-30
: PRIOR APPLICATION NUMBER: JP 2001-272697
: PRIOR FILING DATE: 2001-08-02
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 8425
: LENGTH: 1354
: TYPE: PRT
: ORGANISM: Streptomyces avermitilis
US-10-156-761-8425

```

```

Query Match 5.6%; Score 97; DB 4; Length 1354;
      Beat Local Similarity 24.0%; Pred. No. 6.7;
      Matches 78; Conservative 33; Mismatches 110; Indels 104; Gaps 20;

QY 32 WIAED---HTVALLELDKDRV-----LYATDHSGREALLGGCAVLDFHRV-----72
DB 330 WLEBDGVNHTHEILNKGQVGLTYPPARSSHDTKMLAGGEQALVKITPMFPGSPG 389
QY 73 -----AMAAAGTTANVERFPNNDELHLAST--DSPA--DRV-----TGHRR 111
DB 390 DKGAGAGLAFVDAGAT-----MNNONSPTLYLAASGMDMPVRRBSGIMNHVRLRSYGA 444
QY 112 LRABAILLRTRDRAPFAAPPMDLVE-----SGLRTVTVA--PTVRID--VI 154
DB 445 VTGDP---KDYTKLP--DLPTSAAYTYIVPVANVGSASAKRYVTYLAFFDVRYSRYTV 499
QY 155 ADDMRPELAASAKLTESLRUVSSYHAELFMWTGAFETSGIP--HSLVSAA---ESD 208
DB 500 AGGSAADVTPAPSTYALRLRDPE-----LWMPNGY---GRPLDHLDTLASIDGESD 550
QY 209 RVT-----FGRPFVAVANTDRAPERG--HDSKVLVLTYNERSLRLCGSMLSAVL 259
DB 551 RRTTRFGIRGQGYEYTVPL-----PGGSDRYTQPVLLGVOKARVYRIKC-----L 597
QY 260 LDATMAGLATCTLTHITELHASRDL 284
DB 598 TRATDWGSLWLTSLVVDSTSACTDL 622

```

Search completed: March 23, 2006, 06:28:13  
Job time : 70.6072 secs

Db 6524 RHRDLASIDQVVDVLVLSCEPEADGCPAPEATSSALRRVLEVRVEMLDGARDYTDARLAVL 6583  
Qy 187 T-GAETSEGI PHSLVSAESDRYTFGRDPVYANTDRRPEFGHDSKV---LVYSTY 241  
Db 6584 TRRAVANTSGDVEDDLAAAVRGILRTAQ---ENPRLVVIDHDSOLEVLPVLTGTG 6639  
Qy 242 DNERASLRCGEMLSAVLLDA-----TWAGLATCTTLTHITELH 279  
Db 6640 EPEAA--IRAKGLVLPRLVYKAAVSEKAPAMDAGTVLLTGGTGLGLVA---RLVYTH 6694  
Qy 280 ASRDLVAALIGQATPOLVRYGLAPEME 308  
Db 6695 GARDLVLASRGDTAFGA---VELATELE 6720

RESULT 2  
US-11-072-512-3403  
; Sequence 3403, Application US/11072512  
; Publication No. US20060029945A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHITAKA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOKYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: 064335-0191  
; CURRENT APPLICATION NUMBER: US/11/072, 512  
; CURRENT FILING DATE: 2005-03-07  
; PRIOR APPLICATION NUMBER: US 60/350,978  
; PRIOR FILING DATE: 2002-01-25  
; PRIOR APPLICATION NUMBER: JP 2001-379298  
; PRIOR FILING DATE: 2001-11-05  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3403  
; LENGTH: 534  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-072-512-3403

Query Match 5.5%; Score 95; DB 7; Length 534;  
Best Local Similarity 25.2%; Pred. No. 0.44;  
Matches 41; Conservative 24; Mismatches 52; Indels 46; Gaps 6;  
Qy 187 TGAEETSEGI PHSLVSAESDRYTFGRDPVYANTDRRPEFGHDR-SKVLVSTYDNER 245  
Db 97 SSAVSSSYGYPPSSD-----GR-----ATDQAGFGNDTSLKVGSISSIEGCM 140  
Qy 246 ASLRCGEMLSAV-----LDDATMAGLATCTTLTHITELHASNDVLAALIGQATPOL 298  
Db 141 TGLKIGGLTAAVTKVTGTALSSSGMTSIATNSVPSAAPKPTSWAIAKPKAPQPK 200  
Qy 299 VR-----VGLAPEMEPPP-----ATPRRP 319  
Db 201 LKPKGNVIGSGSAVPPPIKNNMIGTWDKGSVVKAPPTQPV 243

RESULT 3  
US-11-060-005-4

; Sequence 4, Application US/11060005  
; Publication No. US20050260693A1  
; GENERAL INFORMATION:  
; APPLICANT: Irwin H. Gelman  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: A30558-A-FWC-A-A 070156.0597  
; CURRENT APPLICATION NUMBER: US/11/060,005  
; CURRENT FILING DATE: 2005-02-17  
; PRIOR APPLICATION NUMBER: 09/902,432  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 08/978,277  
; PRIOR FILING DATE: 1997-11-25  
; PRIOR APPLICATION NUMBER: 08/665,401  
; PRIOR FILING DATE: 1996-06-18  
; PRIOR APPLICATION NUMBER: 08/635,121  
; PRIOR FILING DATE: 1996-04-19  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 1596  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-11-060-005-4

Query Match 5.3%; Score 90.5; DB 7; Length 1596;  
Best Local Similarity 22.7%; Pred. No. 5.5;  
Matches 62; Conservative 45; Mismatches 123; Indels 43; Gaps 11;  
Qy 65 AYLDHFRVMAAAAGTANVERFP-----NPNDPL-HLASIDFSPADFTVTEGHRRLRADIL 118  
Db 875 AVIDGTR-----AVTSVEERSSPMISASYTEPLEHTAGAMPVEVTEKDIJAETPV 928  
Qy 119 LRRTDRLPFAEPPOWMDVESO-----LRTVTADTVRIDVI-----ADDMREPLAAAS 166  
Db 929 LTGT--LPEGDADHDVTVSEVDTSRAVTAITSEALRTEBEVTEASGAETTDMSAVS 986  
Qy 167 KLTESRLVDSYSAELFWMTGAFETSEGI PHSLVSAESDRYTFGRDPVYANTDR-- 224  
Db 987 QLTOSPTEBAYTVGVV-ESGVLDTEEBERQOQALIQAVADKVEESQVPATQTVGRTG 1045  
Qy 225 -----RPEFGHDSKTVLSTYDNERASLRCGEMLSA---VLDDATMAGLATCTTLHIT 276  
Db 1046 SKALEKVEEVEBDESEVLA---SEKEDVMPKGPVQDGAHEHLAQGETGOATPESLEVP 1101  
Qy 277 ELHASRDLVAALIGQATPOLVRYGLAPEME 309  
Db 1102 EVTADVDVHATC--QVTKLQQLMEQAVAPESSE 1132

RESULT 4  
US-11-060-005-2  
; Sequence 2, Application US/11060005  
; Publication No. US20050260693A1  
; GENERAL INFORMATION:  
; APPLICANT: Irwin H. Gelman  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: A30558-A-FWC-A-A 070156.0597  
; CURRENT APPLICATION NUMBER: US/11/060,005  
; CURRENT FILING DATE: 2005-02-17  
; PRIOR APPLICATION NUMBER: 09/902,432  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 08/978,277  
; PRIOR FILING DATE: 1997-11-25  
; PRIOR APPLICATION NUMBER: 08/665,401  
; PRIOR FILING DATE: 1996-06-18  
; PRIOR APPLICATION NUMBER: 08/635,121  
; PRIOR FILING DATE: 1996-04-19  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 1346  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 06:17:23 ; Search time 7.94316 Seconds  
(without alignments)  
1192.746 Million cell updates/sec

Title: US-10-617-038-25  
Perfect score: 1718  
Sequence: 1 MPDMMTTDVAKSAVQLACR.....PATRRRPIDVEYHVRADHR 331

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 169630 seqs, 28622889 residues

Total number of hits satisfying chosen parameters: 169630

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New:  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/PC1\_NEW\_PUB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101.5	5.9	7102	US-11-143-980-48	Sequence 48, Appl
2	95	5.5	534	US-11-072-512-3403	Sequence 3403, Ap
3	90.5	5.3	1596	US-11-060-005-4	Sequence 4, Appl
4	88.5	5.2	1346	US-11-060-005-2	Sequence 2, Appl
5	84.5	4.9	1734	US-11-192-967-6	Sequence 6, Appl
6	84.5	4.9	1734	US-11-193-715-6	Sequence 6, Appl
7	84	4.9	324	US-11-096-568A-9021	Sequence 9021, Ap
8	84	4.9	358	US-11-096-568A-9020	Sequence 9020, Ap
9	84	4.9	445	US-11-096-568A-9023	Sequence 9023, A
10	81.5	4.7	1264	US-11-096-568A-30287	Sequence 30287, A
11	81.5	4.7	1301	US-11-096-568A-30286	Sequence 30286, A
12	81.5	4.7	1307	US-11-096-568A-30285	Sequence 30285, A
13	81	4.7	311	US-11-000-463-345	Sequence 345, App
14	81	4.7	519	US-10-330-773-519	Sequence 519, App
15	80	4.7	513	US-11-135-667-64	Sequence 64, Appl
16	79.5	4.6	511	US-11-087-099-10103	Sequence 10103, A
17	79	4.6	299	US-10-858-730-17	Sequence 17, Appl
18	78.5	4.6	556	US-11-087-099-3966	Sequence 3966, Ap
19	78	4.6	1680	US-10-517-939-362	Sequence 362, App
20	77.5	4.5	401	US-11-096-568A-17585	Sequence 17585, A
21	77.5	4.5	1308	US-11-124-367A-396	Sequence 396, App
22	77	4.5	310	US-11-096-568A-28488	Sequence 28488, A
23	77	4.5	473	US-11-087-099-7348	Sequence 7348, Ap
24	77	4.5	510	US-11-096-568A-17760	Sequence 17760, A
25	77	4.5	536	US-11-096-568A-17759	Sequence 17759, A

26	77	4.5	1323	US-10-517-939-312	Sequence 312, App
27	77	4.5	2362	US-11-096-568A-31778	Sequence 31778, A
28	77	4.5	2433	US-11-096-568A-31777	Sequence 31777, A
29	77	4.5	2535	US-11-096-568A-31776	Sequence 31776, A
30	76.5	4.5	303	US-11-087-099-9698	Sequence 9698, Ap
31	76.5	4.5	510	US-11-087-099-8796	Sequence 8796, Ap
32	76.5	4.5	937	US-11-017-550-66	Sequence 66, Appl
33	76.5	4.5	1075	US-11-100-640-12	Sequence 12, Appl
34	76.5	4.5	1250	US-11-117-465-62	Sequence 62, Appl
35	76	4.4	351	US-11-072-512-3529	Sequence 3529, Ap
36	76	4.4	412	US-11-096-568A-5868	Sequence 5868, Ap
37	76	4.4	419	US-11-174-150-40	Sequence 40, Appl
38	76	4.4	430	US-11-096-568A-5867	Sequence 5867, Ap
39	76	4.4	431	US-10-525-674-26	Sequence 26, Appl
40	76	4.4	436	US-11-174-150-39	Sequence 39, Appl
41	76	4.4	464	US-11-096-568A-5866	Sequence 5866, Ap
42	76	4.4	523	US-11-087-099-8632	Sequence 8632, Ap
43	76	4.4	673	US-11-096-568A-19860	Sequence 19860, A
44	75.5	4.4	697	US-11-074-176-214	Sequence 214, App
45	75.5	4.4	765	US-11-087-099-7957	Sequence 7957, Ap

ALIGNMENTS

RESULT 1  
US-11-143-980-48  
; Sequence 48, Application US/11143980  
; Publication No. US20050272133A1  
; GENERAL INFORMATION:  
; APPLICANT: He, Min  
; APPLICANT: Hucul, John  
; APPLICANT: Haltli, Bradley A.  
; APPLICANT: Wagneaar, Melissa M.  
; APPLICANT: Graziani, Edmund  
; APPLICANT: Summers, Mia  
; APPLICANT: Kulowski, Kerry  
; APPLICANT: Pong, Kevin  
; TITLE OF INVENTION: Bioynthetic Gene Cluster for the Production of a Complex  
; FILE REFERENCE: AM-101426US  
; CURRENT APPLICATION NUMBER: US/11/143,980  
; CURRENT FILING DATE: 2005-06-03  
; PRIOR APPLICATION NUMBER: US 60/664,483  
; PRIOR FILING DATE: 2005-03-23  
; PRIOR APPLICATION NUMBER: US 60/576,895  
; PRIOR FILING DATE: 2004-06-03  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 48  
; LENGTH: 7102  
; TYPE: PRT  
; ORGANISM: Streptomyces sp.  
US-11-143-980-48

Query Match 5.9%; Score 101.5; DB 7; Length 7102;  
Best Local Similarity 24.7%; Pred. No. 4;  
Matches 96; Conservative 32; Mismatches 122; Indels 139; Gaps 23;

QY 32 WIAEDHTVALFLDMDRLVYATDHSGRALLGCG---AVLDHFRVMAAAGTNAVVERFPN 88  
DB 6359 WIRQSE--VFLLD-----VALPEOVKRDAAKCGHPLALD--NALGIGLGA---FVN 6403  
QY 89 PNDPLHL-----ASIDFSPADFEVTEGHR-----LRADAILAR--- 120  
DB 6404 EPGGAHLHPFMSGCVTLHAGATATVATVLSRAGPTVATIRADITGATVLSIDLAMRPLA 6463  
QY 121 -----RTD---RLPFAEPP-----DWDLY-----ESQLRITVTVA----- 146  
DB 6464 EQRLLAAGSGRQDALFRLBWKELFVPTGATGTPAASGGLGCHDEPRLTALTPAAGVSPQ 6523  
QY 147 ---DTVAIDVIAADM-----RREL--AAASKLTESLR--LYDSY--HAELFW 166





Best Local Similarity 100.0%; Pred. No. 1.2e-108; Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 MPLTTIGQFPAYQLTALIGDLSKVDAKQGDYFTTTTSDHBGKRVVFPMPKDFTFV 60
Db 1 MPLTTIGQFPAYQLTALIGDLSKVDAKQGDYFTTTTSDHBGKRVVFPMPKDFTFV 60
Qy 61 CPTETIAAFSKUNDEFEDRDAQIIGVSIIDSEFAHQWRAQHNDLKTLPFPMISDIKRELSQ 120
Db 61 CPTETIAAFSKUNDEFEDRDAQIIGVSIIDSEFAHQWRAQHNDLKTLPFPMISDIKRELSQ 120
Qy 121 AAGVLANADGVADRYTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQWVRK 180
Db 121 AAGVLANADGVADRYTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQWVRK 180
Qy 181 GDPFLTDAGELLKASA 195
Db 181 GDPFLTDAGELLKASA 195

```

## RESULT 2

AAV43802  
ID AAV43802 standard; protein; 195 AA.

AC AAV43802;  
DT 11-FEB-2000 (first entry)

DE A alkyl hydroperoxide reductase subunit C protein.

XX Hydroperoxide reductase subunit C protein; resistance;  
KM antimicrobial reactive nitrogen intermediate; antibacterial;  
KM nitric oxide; nitrite; nitrosonium; S-nitrosothiol; nitrogen dioxide;  
KM dinitrogen trioxide; dinitrogen tetroxide; septic hypotension; stroke;  
KM infection; vaccine.

XX Mycobacterium tuberculosis.

XX PN W09954479-A2.

XX PD 28-OCT-1999.

XX PF 21-APR-1999; 99WO-US008704.

XX PR 21-APR-1998; 98US-0082573P.

XX PA (CORR ) CORNELL RES FOUND INC.

XX PI Nathan CF, Xie Q, Chen L;

XX DR WPI; 2000-013258/01.

XX DR N-PSDB; AAZ30392.

XX PT Novel uses of alkyl hydroperoxide reductase subunit C (AHRC).

XX PS Claim 40; Page 16-17; 74pp; English.

XX The present sequence represents an alkyl hydroperoxide reductase subunit  
CC C protein, which confers on Mycobacterium tuberculosis resistance against  
CC antimicrobial reactive nitrogen intermediates. The protein is used in a  
CC method for screening therapeutics for their ability to sensitize bacteria  
CC to antibacterial effects of reactive nitrogen intermediates. The method  
CC comprises preparing recombinant host cells lacking endogenous alkyl  
CC hydroperoxide reductase subunit C protein encoding genes, placing cells  
CC in growth medium containing reactive nitrogen intermediates and a test  
CC compound, and determining if the cells survive. The alkyl hydroperoxide  
CC reductase subunit C polypeptides can confer resistance to reactive  
CC nitrogen intermediates, e.g. nitric oxide, nitrite, nitrosonium, S-  
CC nitrosothiol, nitrogen dioxide, dinitrogen trioxide, and dinitrogen  
CC tetroxide. The methods can be used to identify agents that inhibit the  
CC reactive nitrogen intermediate-resistance mechanisms of pathogens. The  
CC alkyl hydroperoxide reductase subunit C polypeptides can also be used to  
CC inhibit production of reactive nitrogen intermediates in the treatment of

CC e.g. septic hypotension, stroke or infections. The alkyl hydroperoxide  
CC reductase subunit C protein or DNA can be used for vaccinating mammals  
CC against infection by M. tuberculosis. The antibodies against the protein  
CC can be used for passively immunising mammals infected with M.  
CC tuberculosis. These products can also be used for detection of M.  
CC tuberculosis

XX SQ Sequence 195 AA;

Query Match 100.0%; Score 1024; DB 3; Length 195;

Best Local Similarity 100.0%; Pred. No. 1.2e-108;

Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 MPLTTIGQFPAYQLTALIGDLSKVDAKQGDYFTTTTSDHBGKRVVFPMPKDFTFV 60
Db 1 MPLTTIGQFPAYQLTALIGDLSKVDAKQGDYFTTTTSDHBGKRVVFPMPKDFTFV 60
Qy 61 CPTETIAAFSKUNDEFEDRDAQIIGVSIIDSEFAHQWRAQHNDLKTLPFPMISDIKRELSQ 120
Db 61 CPTETIAAFSKUNDEFEDRDAQIIGVSIIDSEFAHQWRAQHNDLKTLPFPMISDIKRELSQ 120
Qy 121 AAGVLANADGVADRYTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQWVRK 180
Db 121 AAGVLANADGVADRYTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQWVRK 180
Qy 181 GDPFLTDAGELLKASA 195
Db 181 GDPFLTDAGELLKASA 195

```

## RESULT 3

AAM50749  
ID AAM50749 standard; protein; 195 AA.

AC AAM50749;

DT 18-APR-2002 (first entry)

DE Mycobacterium tuberculosis immunodominant Mtb protein APhc2.

XX Mtb; APhc2; immunogen; mycobacteria; immunisation; vaccine.

XX OS Mycobacterium tuberculosis.

XX PN W0200204018-A2.

XX PD 17-JAN-2002.

XX PF 10-JUL-2001; 2001WO-US021717.

XX PR 10-JUL-2000; 2000US-0217646P.

XX PA (COLS ) UNIV COLORADO STATE RES FOUND.

XX PI Orme IM, Belisle JT;

XX DR WPI; 2002-164602/21.

XX The present sequence is that of the Mycobacterium tuberculosis (Mtb)  
CC strain H37Rv cytosolic Rv2428 gene product, designated APhc2. This is one  
CC of 31 immunodominant secreted or cytosolic Mtb proteins of strain H37Rv  
CC (see AAM50729-59) discovered through the use of 2-dimensional liquid  
CC phase electrophoresis coupled with an in vitro interferon-gamma assay and  
CC liquid chromatography-mass spectrometry. The immunogens stimulate a  
CC strong interferon-gamma response from T cells of M. tuberculosis infected  
CC mice. The invention provides vaccine compositions for boosting immunity  
CC to mycobacteria when administered in mid-life to a subject who has been

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 04:56:29 ; Search time 51.844 Seconds  
(without alignments)  
1652.628 Million cell updates/sec

Title: US-10-617-038-26

Perfect score: 1024  
Sequence: 1 MPLLTIGQFPAYQLTALIG.....CMNRKGDPTLDAGELTKRSA 195

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: A\_Geneseq\_21:\*  
2: geneseqp1980s:\*  
3: geneseqp1990s:\*  
4: geneseqp2000s:\*  
5: geneseqp2001s:\*  
6: geneseqp2002s:\*  
7: geneseqp2003as:\*  
8: geneseqp2003bs:\*  
9: geneseqp2004s:\*  
9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1024	100.0	195	2	AAR98369 Mycobacte
2	1024	100.0	195	3	AAY43802 A alkyl h
3	1024	100.0	195	5	AAM50749 Mycobacte
4	1024	100.0	195	6	ABU34818 Protein e
5	1024	100.0	195	6	ABU36784 Protein e
6	1024	100.0	195	6	AD137305 M. tuberc
7	935	91.3	195	6	ABU33962 Protein e
8	926	90.4	195	6	AA667761 Amino aci
9	904	88.3	195	6	ABU36067 Protein e
10	901.5	88.0	193	2	AA14303 M.avium a
11	898	87.7	195	7	ADB74335 Mycobacte
12	693	67.7	198	6	ABU25828 Protein e
13	566.5	55.3	179	9	ABE39863 L. pneumo
14	566.5	55.3	198	9	ABE36453 L. pneumo
15	530.5	51.8	188	8	ADS30355 Bacteri
16	524.5	51.2	184	9	ABW91466 M. xanthu
17	432.5	42.2	182	6	ABU21078 Protein e
18	410.5	40.1	182	6	ABU22889 Protein e
19	361.5	35.3	180	8	AD28683 Bacteri
20	351.5	34.5	193	8	ADS42506 Bacteri
21	351.5	34.3	192	8	ADS29794 Bacteri
22	351	34.3	199	9	AEA00256 Leishmani
23	350.5	34.2	200	8	ADM48264 Polypepti
24	350.5	34.2	200	8	ADN20134 Bacteri

25	350.5	34.2	283	5	AAE29736	AAE29736 Mycobacte
26	350.5	34.2	282	5	AAE29727	AAE29727 Leishmani
27	350.5	34.2	290	5	AAE29725	AAE29725 Leishmani
28	350.5	34.2	321	5	AAE29728	AAE29728 Leishmani
29	350.5	34.2	328	5	AAE29726	AAE29726 Leishmani
30	350.5	34.2	446	6	ABP55386	ABP55386 Human col
31	350.5	34.2	926	7	ADB78891	ADB78891 Leishmani
32	350.5	34.2	930	5	AAE29731	AAE29731 Mycobacte
33	350.5	34.2	930	7	ADA26364	ADA26364 Mycobacte
34	350.5	34.2	955	7	ADB78890	ADB78890 Leishmani
35	350.5	34.2	982	5	AAU71855	AAU71855 Leishmani
36	350.5	34.2	982	5	AAU71309	AAU71309 MAPS1A-M1
37	350.5	34.2	982	7	ADB78858	ADB78858 Leishmani
38	350.5	34.2	1427	5	AAU71857	AAU71857 Leishmani
39	350.5	34.2	1427	5	AAU71311	AAU71311 MAPS1A-M1
40	350.5	34.2	1427	7	ADB78860	ADB78860 Leishmani
41	350.5	34.2	1464	6	ABR54576	ABR54576 Protease
42	350.5	34.2	1464	7	ADB14458	ADB14458 Expressio
43	350.5	34.2	1464	7	ADG26981	ADG26981 Human pro
44	350.5	34.2	1641	5	AAU71856	AAU71856 Leishmani
45	350.5	34.2	1641	5	AAU71310	AAU71310 MAPS1A-M1

#### ALIGNMENTS

RESULT 1	
AA98369	AA98369 standard; protein; 195 AA.
XX	
XX	AA98369;
AC	
XX	16-OCT-2003 (revised)
DT	19-SEP-1996 (first entry)
DT	
XX	
DE	Mycobacterial Ahpc polypeptide.
XX	
XX	ahpcF operon; Ahpc, AhpF, alkyl hydroperoxide reductase; isoniazid;
KM	antibiotic resistance; mycobacteria; therapy; diagnosis; vaccine.
KM	
XX	
OS	Mycobacterium bovis; strain ATCC 35729.
XX	
PN	W09619578-A2.
XX	
PD	27-JUN-1996.
XX	
PF	20-DEC-1995; 95WO-NZ000133.
XX	
PR	20-DEC-1994; 94NZ-00270222.
XX	
PA	(NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.
XX	
PI	Collins DW, Wilson T;
XX	
DR	WPI; 1996-309592/31.
XX	N-PSDB; AATJ1461.
PT	Polynucleotide(s) derived from the mycobacteria ahpcF operon - useful in
PT	diagnosis and treatment of mycobacterial diseases.
XX	
PS	Example 2; Fig 2; 55pp; English.
XX	
CC	The ahpcF operon (AATJ1461) of isoniazid (INH)-resistant Mycobacterium
CC	bovis ATCC encodes polypeptides Ahpc (AA98369) and AhpF (AA98370),
CC	which likely combine to form an active alkyl hydroperoxide reductase that
CC	may be either a direct target for INH or acts to confer INH resistance.
CC	Isolated Ahpc and/or AhpF polypeptides can be prepd. e.g. in mammalian or
CC	yeast systems. They are useful for drug screening, in immunosays and in
CC	vaccine prodn. (updated on 16-OCT-2003 to standardise OS field)
XX	
SQ	Sequence 195 AA;
Query Match	100.0%; Score 1024; DB 2; Length 195;

Query Match 4.7%; Score 81; DB 7; Length 311;  
Best Local Similarity 21.8%; Pred. No. 4.5;  
Matches 69; Conservative 42; Mismatches 118; Indels 88; Gaps 16;

```
QY 13 SAVOLACRAPSLHNSQPMRWIAEDHTVALFLDKDRLVYATDHSREALLGCAVLDPHFRV 72
DQ 31 TSVEICRSLDFQATWTF-WYRQ-----FPKQSLMIMATSNESKATVEQGVKDKFLI 83
QY 73 AMAAGTTANVERFPNPD-PLHLASIDFSPAD-----FVTEGRLADALLARTDRLP 126
DQ 84 NHAALTJLTLTVAHAPEDSFYICSAESTSDPKNQFEGPGRLLTV-----TEDLK 136
QY 127 FAEPD---MDLVESQRTTVAIDTVRIDVIADDMPELAASKLITSLRLYDSSYAE 183
DQ 137 NVPEPEVAVPEPSEAEISHTOKATLV---CLATGFPD-----HVEL 175
QY 184 FWTGAFETSEGI---PHSLVSAESD-RVTFGRDPFVANTDRP----- 226
DQ 176 SMVNGKEVHSGVSTDPQLKEQPALNDSRYCLSSRLVSATFWQNPNNHPRCQVQFYGL 235
QY 227 -----EFGHDSKVL--VLSTYDNERASILRCG-----EMLSAVLLDATMAGLATCTL 272
DQ 236 SENDEWQDRACKPTQIVSABAMGRAD---CGFTSESYQGVLSATILYEILGKAT--- 289
QY 273 THITELHASRDLVAALI 289
DQ 290 -----LYAV-LVSALV 299
```

## RESULT 14

```
US-10-330-773-519
; Sequence 519, Application US/10330773
; Publication No. US20060040262A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001300
; CURRENT APPLICATION NUMBER: US/10/330,773
; CURRENT FILING DATE: 2002-12-27
; NUMBER OF SEQ ID NOS: 981
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 519
; LENGTH: 519
; TYPE: PRF
; ORGANISM: Mus musculus
US-10-330-773-519
```

Query Match 4.7%; Score 81; DB 6; Length 519;

Best Local Similarity 20.5%; Pred. No. 9.2;  
Matches 53; Conservative 27; Mismatches 77; Indels 102; Gaps 12;

```
QY 9 DVKSAVO-LACRAPSLHNSQPMRWIAEDHTVALFLDKDRLVY-----ATDHSGR 57
DQ 166 DILISVIEKMKCNLP--NQOQPERWDEDAISRLGTGAEWTFYHIKQEPSSASHTGC 223
QY 58 EALLGCCGAVLDHFRVMAAAAGTTANVERFPNPDPLHLASIDFSPADFTGSHLRDAI 117
DQ 224 E---GCAA-----LVGGEVEYAEEDFSDVDAAR--EV 251
QY 118 LARTRDLFAEPDMDLVESQRTTVAIDTVR-IDVIAD-----MRPELAASK 167
DQ 252 MLK--SRVPGE--DWVLPQCITLTVHPPIKDIAYVAQNFCACGCTPIQPKFVKLR 307
QY 168 LTESLRLY-----DSSYHAELFWWTGAFB 191
DQ 308 YCEYLGYFASCHSAESCIPIRLTMWDRKXQVSDPSKMLDVSWHQGVFKLLGSH-- 366
QY 192 TSEGIPIHSLVSAESDRV 210
DQ 367 -----HSLYAKAKELDRV 379
```

Query Match 4.7%; Score 80; DB 7; Length 513;  
Best Local Similarity 22.1%; Pred. No. 11;  
Matches 81; Conservative 36; Mismatches 95; Indels 154; Gaps 21;

```
QY 6 VTTDVKSAVOLACRAPSLHNSQPMR-----WIAEDHTVALFLDKDRLVYATDHSGR--- 57
DQ 229 LITEVMTST--ASCAPAH--EPVRPHQOAVSAHINRLTGSGVLDHNRDCDGRVOD 284
QY 58 EALLGC-----GAV---LDHFRVMAAAAGTTANVERFPNPDPLHLASIDFSPADFTG 109
DQ 285 AVSIRCLPQVHGAVRDALDHLRAAVATELNSAT-----DNPL-----VPPSGTVVSG 331
QY 110 HRRLADAILY-----RRTRDL-----PFAEPDMDLVESQRT 142
DQ 332 GNTHGEVTLARLYAASALAEIAISBRRTDRLNPTQEPLEPFLAP-----DSGLHS 365
QY 143 TVTADTVRIDVIADDMKPELAASKLITSLRLYDSSYHAELFWWTGAFETSEGI 199
DQ 386 GL-----MLPQYTAAS-LVNDLR-----SLGQPATLDN 412
QY 200 SLVSAESDRVTFGRDPFVANTDRRPERGHSKVLVSTYD-----NERASL-----LR 250
DQ 413 ASVSGAQEDHVSMSAG-----AAVNFREAVEKATVVGVELL 449
QY 251 CGEMLSAVLLDATMAGLATCTLTHITELHASRDLVAALIGOPATPOA-----LVR 300
DQ 450 CGAQ-GREFLDPLALAGTA-----AAVLDVRSVSEPPAGDRALADMAAVGDDLV 499
QY 301 VGLAPE 306
DQ 500 AGLVED 505
```

Search completed: March 23, 2006, 06:29:36  
Job time: 8.94316 secs







PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362899P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR N-PSDB; ACA40654.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 25; SEQ ID NO 64708; 1766bp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 195 AA;  
Query Match 100.0%; Score 1024; DB 6; Length 195;  
Best Local Similarity 100.0%; Pred. No. 1.2e-108;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MPLLITGDOPPAVQLTALIGDLSKVDAKQPGDYFTTTSDEHPGKRWVFPKQDTFV 60  
DB 1 MPLLITGDOPPAVQLTALIGDLSKVDAKQPGDYFTTTSDEHPGKRWVFPKQDTFV 60  
QY CPTETIAAFSKLNDEFEDRDAQILGVSTIDSEFAHFQWRQNHDKLTLPFPMLSDIKRELSQ 120  
DB 61 CPTETIAAFSKLNDEFEDRDAQILGVSTIDSEFAHFQWRQNHDKLTLPFPMLSDIKRELSQ 120  
QY 121 AAGVLANADVADRTFTVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDCLCACMWRK 180  
DB 121 AAGVLANADVADRTFTVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDCLCACMWRK 180  
QY 181 GDPTLDAGEELIKASA 195  
DB 181 GDPTLDAGEELIKASA 195  
RESULT 6

AD137305  
ID AD137305 standard; protein: 195 AA.  
XX  
XX AD137305;  
AC  
XX 22-APR-2004 (first entry)  
DT  
XX  
XX M. tuberculosis low oxygen induced antigen Rv2428 SEQ ID NO:26.  
DE  
XX mycobacterial infection; vaccine; tuberculosis;  
KW Mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;  
XX low oxygen induced antigen.  
XX  
XX Mycobacterium tuberculosis.  
XX  
XX WO2004006952-A2.  
XX  
XX 22-JAN-2004.  
XX  
XX 08-JUL-2003; 2003WO-DK000477.  
XX  
XX 13-JUL-2002; 2002DK-00001098.  
XX  
XX (STAT-) STATENS SERUM INST.  
XX  
XX Andersen P, Rosenkrands I, Stryhn A;  
XX WPI; 2004-122778/12.  
DR N-PSDB; AD137350.  
XX  
XX Claim 3; SEQ ID NO 26; 76pp; English.  
XX  
XX The present invention describes polypeptides or their fragments, which  
CC are expressed during the latent stage of a mycobacterial infection,  
CC and/or nucleic acids encoding the polypeptides, which are useful for  
CC creating a therapeutic vaccine against tuberculosis. Also described: (1)  
CC a therapeutic vaccine against tuberculosis comprising one or more  
CC polypeptides; (2) a method for treating an animal, including a human  
CC being, with tuberculosis caused by virulent mycobacteria, e.g. by  
CC Mycobacterium tuberculosis, *M. africanum* or *M. bovis*; (3) a method for  
CC immunising an animal, including a human being, against tuberculosis  
CC caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis  
CC caused by virulent mycobacteria in an animal, including a human being;  
CC (5) a method for diagnosing previous or ongoing infection with a virulent  
CC mycobacterium; and (6) a method of diagnosing Mycobacterium tuberculosis  
CC infection in a subject. The polypeptides have antibacterial activities,  
CC and can be used in vaccines and in gene therapy. The polypeptides are  
CC useful for the manufacture of a therapeutic vaccine for treating an  
CC individual who is infected by a virulent mycobacterium, e.g. *M.*  
CC tuberculosis, and who is not vaccinated with BCG against tuberculosis.  
CC The present sequence represents a low oxygen induced antigen, which is  
CC used in the exemplification of the present invention.  
XX  
SQ Sequence 195 AA;  
Query Match 100.0%; Score 1024; DB 8; Length 195;  
Best Local Similarity 100.0%; Pred. No. 1.2e-108;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MPLLITGDOPPAVQLTALIGDLSKVDAKQPGDYFTTTSDEHPGKRWVFPKQDTFV 60  
DB 1 MPLLITGDOPPAVQLTALIGDLSKVDAKQPGDYFTTTSDEHPGKRWVFPKQDTFV 60  
QY CPTETIAAFSKLNDEFEDRDAQILGVSTIDSEFAHFQWRQNHDKLTLPFPMLSDIKRELSQ 120  
DB 61 CPTETIAAFSKLNDEFEDRDAQILGVSTIDSEFAHFQWRQNHDKLTLPFPMLSDIKRELSQ 120  
QY 121 AAGVLANADVADRTFTVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDCLCACMWRK 180  
DB 121 AAGVLANADVADRTFTVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDCLCACMWRK 180

Db 121 AAGVLANADGVADRVTFIVDPNNIEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQNRK 180  
QY 181 GDDPTLDAGELTKASA 195  
Db 181 GDDPTLDAGELTKASA 195

RESULT 7  
ABU33962 standard; protein; 195 AA.

ABU33962;  
19-JUN-2003 (first entry)  
Protein encoded by Prokaryotic essential gene #19489.

XX Anti-sense; prokaryotic essential gene; cell proliferation; drug design.  
XX Mycobacterium avium.

OS WO200277183-A2.  
XX 03-OCT-2002.

PD 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.  
XX 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.

XX (EliT-) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Hasebeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX

DR WPI; 2003-029926/02.  
XX N-PSDB; ACN37832.

PT New anti-sense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 61886; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 anti-sense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the anti-sense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC anti-sense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The anti-sense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pcc\_sequences

XX Sequence 195 AA;  
Query Match 91.3%; Score 935; DB 6; Length 195;  
Best Local Similarity 90.3%; Pred. No. 2e-98;  
Matches 176; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 1 MRLTTIGDQFPAYQYLTALIGDLSKVDAKQDGYFTTTSDEHPGKRVFFWPKDFTFV 60  
DB 1 MRLTTIGDQFPAYELTALIGDLSKVDAKQDGYFTTTSDEHAGKRVFFWPKDFTFV 60  
QY 61 CPTETAAFSKUNDEFEDBDAGILGVSTIDSEFAHQMAQHNDLTPPPMISDIKRELSQ 120  
DB 61 CPTETATFGKLNDEFEDBDAGVIGVSTIDSEFAHQMAQHNDLTPPPMISDIKRELSL 120  
QY 121 AAGVLANADGVADRVTFIVDPNNIEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQNRK 180  
DB 121 AAGVLANADGVADRVTFIVDPNNIEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQNRK 180  
QY 181 GDDPTLDAGELTKASA 195  
DB 181 GDDPTLDAGELTKASA 195

QY 181 GDDPTLDAGELTKASA 195  
DB 181 GDDPTLDAGELTKASA 195

RESULT 8  
AAB67761  
ID AAB67761 standard; protein; 195 AA.

AC AAB67761;  
XX 11-JUN-2001 (first entry)  
XX Amino acid sequence of the IO1 antigen.

DE IO1 antigen; paratuberculosis; IO2 antigen; Crohn's disease; vaccine.

XX Mycobacterium avium.  
OS Key Location/Qualifiers  
FH Misc-difference 192  
FT /note="encoded by AAG"

XX WO200118212-A2.

XX 15-MAR-2001.

XX 06-SEP-2000; 2000WO-GB003430.

XX 06-SEP-1999; 99NO-00004317.

XX (VESO-) VESO AS.

XX (JONE/) JONES E J.

XX Oleen I, Wiker H;

XX WPI; 2001-244574/25.

XX N-PSDB; AAF80135.

PT Diagnosing or monitoring infection by *Mycobacterium avium* spp.  
PT paratuberculosis, by assessing presence or concentration of the bacterial  
PT derived antigens or antibodies cross-reacting with the antigens.

XX Claim 1; Fig 1; 61pp; English.

XX The present sequence represents the IO1 antigen of *Mycobacterium avium*.  
CC The specification describes a method for diagnosing and monitoring  
CC paratuberculosis infection in a human or non-human animal. The method

CC comprises assessing the presence or concentration of a polypeptide  
 CC derived from paratuberculosis, a nucleic acid encoding the polypeptide,  
 CC antibodies directed to the polypeptide or one or more infection markers  
 CC which result from in vitro presentation of the polypeptide in a sample  
 CC from the human or non-human animal. The polypeptide is especially the I01  
 CC or I02 antigen. I01 and I02 polypeptide, polymucleotides, and antibodies  
 CC are useful for assessing the presence or concentration of  
 CC paratuberculosis in a sample. They are useful for stimulating an immune  
 CC response against or for preventing infection by *M. avium* ssp  
 CC paratuberculosis, for diagnosing and monitoring *M. avium* ssp  
 CC paratuberculosis infection and for preventing paratuberculosis or Crohn's  
 CC disease in a human or non-human animal, especially a human or a cow  
 XX  
 SQ Sequence 195 AA;  
 Query Match 90.4%; Score 926; DB 4; Length 195;  
 Best Local Similarity 89.2%; Pred. No. 2.1e-97;  
 Matches 174; Conservative 7; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 MFLTTIGDQPPAYQLTALIGDLSKVDAKQPGDYFTTITSDEHFGKRWVFPPKDPFTFV 60  
 DB 1 MFLTTIGDQPPAYQLTALIGDLSKVDAKQPGDYFTTITSDEHFGKRWVFPPKDPFTFV 60  
 QY 61 CPTETIAAFSKLNDPEFEDRDAGIIGVSIIDSEFAHFQMRQAQNDLKTLPFPMLSDIKRELISQ 120  
 DB 61 CPTETIAAFSKLNDPEFEDRDAGIIGVSIIDSEFAHFQMRQAQNDLKTLPFPMLSDIKRELISL 120  
 QY 121 AAGVLANADVADRVTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQMRK 180  
 DB 121 AAGVLANADVADRVTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQMRK 180  
 QY 181 GDPTLNDAGELLKASA 195  
 DB 181 GDPTLNDAGELLKASA 195  
 RESULT 9  
 ABU36067  
 ID ABU36067 standard; protein, 195 AA.  
 XX  
 AC ABU36067;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #21594.  
 XX  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 OS  
 XX Mycobacterium leprae.  
 PN  
 XX WO200277183-A2.  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Mang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR N-PSDB; ACA39937.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.

XX  
 PS Claim 25; SEQ ID NO 63991; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 195 AA;  
 Query Match 88.3%; Score 904; DB 6; Length 195;  
 Best Local Similarity 86.7%; Pred. No. 7.2e-95;  
 Matches 169; Conservative 12; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 MFLTTIGDQPPAYQLTALIGDLSKVDAKQPGDYFTTITSDEHFGKRWVFPPKDPFTFV 60  
 DB 1 MSLTIGQPPAYQLTALIGDLSKVDAKQPGDYFTTIVSSDSHPGKRWVFPPKDPFTFV 60  
 QY 61 CPTETIAAFSKLNDPEFEDRDAGIIGVSIIDSEFAHFQMRQAQNDLKTLPFPMLSDIKRELISQ 120  
 DB 61 CPTETIAAFSKLNDPEFEDRDAGIIGVSIIDSEFAHFQMRQAQNDLKTLPFPMLSDIKRELISA 120  
 QY 121 AAGVLANADVADRVTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQMRK 180  
 DB 121 AAGVLANADVADRVTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQMRK 180  
 QY 181 GDPTLNDAGELLKASA 195  
 DB 181 GDPTLNDAGELLKASA 195  
 RESULT 10  
 AAR14303  
 ID AAR14303 standard; protein, 193 AA.  
 XX  
 AC AAR14303;  
 XX  
 DT 02-JAN-1992 (first entry)  
 XX  
 DE *M. avium* av1-3 protein.  
 XX  
 KW mycobacterial infection; diagnostic kit.  
 OS  
 XX Mycobacterium avium.  
 PN JP03206098-A.

XX 09-SEP-1991.  
PD 29-DEC-1989; 89JP-00344183.  
XX 29-DEC-1989; 89JP-00344183.  
XX 29-DEC-1989; 89JP-00344183.  
XX (AJIN ) AJINOMOTO KK.  
PA WPI; 1991-307344/42.  
DR N-PSDB; ANQ14175.  
XX Mycobacterium avium derived AVI-3 protein - prep'd. from new transformant  
PT transformed by plasmid, used as diagnostic agent for mycobacterium avium  
XX infection.  
PS Claim 1; Page 1 and Fig 3; 12pp; Japanese.  
XX The sequence coding for the avi-3 protein is also claimed. Transformants  
CC containing the coding sequence are cultured to produce recombinant avi-3  
CC for diagnosis of M.avium infection  
XX SQ . Sequence 193 AA;

Query Match 88.0%; Score 901.5; DB 2; Length 193;  
Best Local Similarity 88.1%; Pred. No. 1.4e-94;  
Matches 171; Conservative 6; Mismatches 16; Indels 1; Gaps 1;

QY 2 PLVTIGDOPFAYVQLTALIGDLSKVDAKOPGDYFTTTSDEHFGKRVVFFWPKDFTFVC 61  
Db 1 PLVTIGDOPFAYVQLTALIGDLSKVDAKOPGDYFTTTSDEHFGKRVVFFWPKDFTFVC 60  
QY 62 PTEIAAFSKLNDEFEDRDQAQILGVSIDSEFAHFQWRAQHNDLKTLPPMLSDIKRELSSQA 121  
Db 61 PTEIAAFSKLNDEFEDRDQAQILGVSIDSEFAHFQWRAQHNDLKTLPPMLSDIKRELSSQA 120  
QY 122 AGVINADGVADRVTFTVDPNNIEQFVSATAGSVGRNDEVLRVLDALQSDDELCAQNRKXG 181  
Db 121 TGVINAD-AGDRATFTVDPNNIEQFVSATAGSVGRNDEVLRVLDALQSDDELCAQNRKXG 179  
QY 182 DPTLDAGELTKASA 195  
Db 180 DPTLNATELLTKASA 193

RESULT 11  
ADB74335  
ID ADB74335 standard; protein; 195 AA.

XX ADB74335;  
XX 04-DEC-2003 (first entry)  
XX Mycobacterium leprae non-naturally occurring peptide #58.  
XX Non-naturally occurring peptide; anion pump protein; tuberculosis;  
KM hypersensitivity reaction; tuberculostatic.  
XX Mycobacterium leprae.  
XX US6583266-B1.  
XX 24-JUN-2003.  
XX 16-SEP-1994; 94US-00311731.  
XX 19-AUG-1993; 93US-00109181.  
XX 22-OCT-1993; 93US-00142558.  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX Smith DR, Mao J;  
XX

DR WPI; 2003-656441/62.  
DR N-PSDB; ADB74276.  
XX New Mycobacterium tuberculosis anion pump peptide useful for as  
PT tuberculosis vaccine and diagnosis of tuberculosis infection.  
XX Disclosure; SEQ ID NO 84; 26pp; English.

XX The invention relates to a non-naturally occurring peptide of  
CC Mycobacterium tuberculosis comprising an amino acid sequence  
CC corresponding to an anion pump protein. The invention also relates to a  
CC non-naturally occurring nucleic acid corresponding to a DNA sequence of  
CC Mycobacterium tuberculosis or Mycobacterium leprae. The new peptide is  
CC useful as a vaccine against Mycobacterium tuberculosis or Mycobacterium  
CC leprae or for screening for new tuberculosis drugs. Purified proteins  
CC derived from the sequences of the invention may elicit a specific immune  
CC response. The peptide may also be used to detect hypersensitivity  
CC reactions of individuals exposed to Mycobacterium tuberculosis or  
CC Mycobacterium leprae. The proteins and peptides may be affixed to solid  
CC supports to detect antibodies typical of hypersensitivity reactions, from  
CC a patient's sera. This sequence represents Mycobacterium leprae non-  
CC naturally occurring peptide of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html.

SQ Sequence 195 AA;

Query Match 87.7%; Score 898; DB 7; Length 195;  
Best Local Similarity 85.6%; Pred. No. 3.5e-94;  
Matches 167; Conservative 14; Mismatches 14; Indels 0; Gaps 0;

QY 1 MPLVTIGDOPFAYVQLTALIGDLSKVDAKOPGDYFTTTSDEHFGKRVVFFWPKDFTFV 60  
Db 1 MSLSLISQGFAYVQLTALIGDLSKVDAKOPGDYFTTTSDEHFGKRVVFFWPKDFTFV 60  
QY 61 CPTLIAAFSKLNDEFEDRDQAQILGVSIDSEFAHFQWRAQHNDLKTLPPMLSDIKRELSSQA 120  
Db 61 CPTLIAAFSKLNDEFEDRDQAQILGVSIDSEFAHFQWRAQHNDLKTLPPMLSDIKRELSSQA 120  
QY 121 AAGVLNADGVADRVTFTVDPNNIEQFVSATAGSVGRNDEVLRVLDALQSDDELCAQNRKXG 180  
Db 121 ASGVLNADGVADRVTFTVDPNNIEQFVSATAGSVGRNDEVLRVLDALQSDDELCAQNRKXG 180  
QY 181 GDPTLDAGELTKASA 195  
Db 181 GDPTLNATELLTKASA 195

RESULT 12  
ABU25828  
ID ABU25828 standard; protein; 198 AA.

XX ABU25828;  
XX 19-JUN-2003 (first entry)  
XX Protein encoded by Prokaryotic essential gene #11355.  
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX Corynebacterium diphtheriae.  
XX WO200277183-A2.  
XX 03-OCT-2002.  
XX 21-MAR-2002; 2002WO-US009107.  
XX 21-MAR-2001; 2001US-00815242.  
XX 06-SEP-2001; 2001US-00948993.  
XX 25-OCT-2001; 2001US-0342923P.  
XX 08-FEB-2002; 2002US-00072851.  
XX

PR 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JM;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR N-PSDB; ACA29698.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX  
PS Claim 25; SEQ ID NO 53752; 1766bp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pcr\_sequences  
XX  
SQ Sequence 198 AA;  
Query Match 67.7%; Score 693; DB 6; Length 198;  
Best Local Similarity 67.9%; Pred. No. 1.2e-70;  
Matches 125; Conservative 29; Mismatches 30; Indels 0; Gaps 0;  
QY 1 MPLLITDQPPAYVLTALIGDLSKVDKOPGDFYTTITSDEHSGKRVVFPWXPDTFV 60  
DB 1 MSILTVKEKPEFNLTKGGDLHDVNASQEDYFEVLSDKYGGKKRVFFYKDPFV 60  
QY CPEEIAAFSKLNDFFEDDAQILGVSIDSEFAHFQWRAQHNDLKTLPFPMISDIKRELSQ 120  
DB 61 CPEEIAAFGLKDEFFQDRDQILGSGSIDNEFSHPMAKTHBELKTVFPLFSIDKHLILK 120  
QY 121 AAGVLNAGVADRVTFTVPDNEIQFVSATAGSVGRVNDVLTALDLSDELCAQWVRK 180  
DB 121 ALGVENNEGVADRATFTIIDPGIIGFVSVTPDAVGRVNDVLTALDLSDELCAQWVRK 180  
QY 181 GDPF 184  
DB 181 NDPT 184

RESULT 13  
AEB39863

ID AEB39863 standard; protein; 179 AA.  
XX  
XX AEB39863;  
AC  
XX 08-SEP-2005 (first entry)  
DT  
XX L. pneumonia protein SEQ ID NO 4195.  
DE  
XX detection; infection; Antibacterial; Vaccine.  
KM  
XX Legionella pneumonia.  
OS  
XX W02005049642-A2.  
PN  
XX  
PD 02-JUN-2005.  
PD  
XX  
PF 23-SEP-2004; 2004WO-1B003578.  
XX  
XX 21-NOV-2003; 2003FR-00013687.  
PR  
XX (INSP ) INST PASTEUR.  
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
PA (UPLY-) UNIV LYON 1 BERNARD CLAUDE.  
PA (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
PI Buchrieser C, Tichit M, Etienne J, Ma L, Cazalec C, Glaser P;  
PI Rumsloek C, Bouchier C, Zidane N, Magnier A, Kunat F, Vandenech F;  
PI Jarraud S;  
XX WPI; 2005-388305/40.  
DR  
XX  
XX New genome of *Legionella pneumoniae* Paris strain and derived  
PT polypeptides, useful for detection or identification of the strain and  
PT for treatment and prevention of infections.  
PS  
XX Claim 3; SEQ ID NO 4195; 660bp; English.  
XX  
XX The invention relates to an isolated or purified nucleotide sequences (1)  
CC from *Legionella pneumoniae* Paris strain. (I), and their related  
CC sequences or fragments, are useful as primers and probes for detection  
CC and amplification, including differentiation between the Paris and  
CC Philadelphia strains of *Legionella pneumoniae* and to prepare recombinant  
CC (hybrid) polypeptides (II). (II) are also useful for preparation of  
CC specific antibodies (Ab), also used for detection/identification of  
CC *Legionella*, and some (I), specifically those involved in synthesis of  
CC surface proteins, are targets for identification of inhibitors. (II), or  
CC vectors that contain (I), are useful as vaccines and immunogenic  
CC compositions, for treatment and prevention of infections by *L.*  
CC *pneumophila*. The present sequence represents the amino acid sequence of a  
CC *L. pneumoniae* protein.  
XX  
SQ Sequence 179 AA;  
Query Match 55.3%; Score 566.5; DB 9; Length 179;  
Best Local Similarity 58.2%; Pred. No. 3.5e-56;  
Matches 107; Conservative 26; Mismatches 42; Indels 9; Gaps 2;  
QY 3 LTTIGDQPPAYVLTALIGDLSKVDKOPGDFYTTITSDEHSGKRVVFPWXPDTFVCP 62  
DB 1 MITVGNKPEFQKATVANDINNA-----FQMSNERYRKMVLVFPWXPDTFVCP 52  
QY 63 TEIAAFSKLNDFFEDDAQILGVSIDSEFAHFQWRAQHNDLKTLPFPMISDIKRELSQA 122  
DB 53 TEIAAFGRILNSFADDAQILGSGTSEFVHLMRQHNPDLKLPFPMADVGRBELTANL 112  
QY 123 GVL-NADGVADRVTFTVPDNEIQFVSATAGSVGRVNDVLTALDLSDELCAQWVRK 181  
DB 113 GILDEBEGAQRATFTVDPGIGITRFVWVNDLVNGRPAEVLTVLADLSDELCPQWVRK 172  
QY 182 DPTL 185  
DB 173 ERTI 176



Query Match	55.3%	Score 566.5	DB 9	Length 198
Best Local Similarity	58.2%	Pred. No. 4e-56		
Matches 107	Conservative 26	Mismatches 42	Indels 9	Gaps 2
3	LTITGDPPAYQLTALIGDLSKVADKQGDYFTTITSDEHPEKMRVVFPMKDTFTVCP	62		
20	MITVNCKEPPEQLKATVANDINNA-----FQMSINERYRQKMLVVFPMKDTFTVCP	71		
63	TEIAAFSKINDSEFEDRDAQILGVISISEPAHFQWRQANDLKTLPFPMISDIRSELSQA	122		
72	TEIAEFGRLNSEFAIRDAQILGSGTSEFVHLAMRQHPDLKDLPPMLADVRELTANT	131		
123	GVLT-NADGVADVTFIVDPNNELQFVSATAGSVGRVNDVLTARLADLQSDLCACPMRK	181		
132	GLIDDEEGVAQRAITFIVDPQGITRFVMTVDLNGRPAEVLRLVLDLQTDDELCPCKMK	191		
182	DPFL 185			

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 05:11:18 ; Search time 7.75813 Seconds  
(without alignments)  
2418.401 Million cell updates/sec

Title: US-10-617-038-26

Perfect score: 1024

Sequence: 1 MPLLTTGDPFPAYQLTALIG.....CNMRKGDPTLDAGELLKASA 195

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1.\*  
2: p1r2.\*  
3: p1r3.\*  
4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1024	100.0	195	2 S71013	alkyl hydroperoxid
2	1024	100.0	195	2 B70679	alkyl hydroperoxid
3	935	91.3	195	2 A43858	alkyl hydroperoxid
4	904	88.3	195	2 B87164	alkyl hydroperoxid
5	444	43.4	184	2 A43581	alkyl hydroperoxid
6	361.5	35.3	180	2 B69867	2-cys peroxidorexi
7	360	35.2	199	2 S73193	hypothetical prote
8	351.5	34.3	203	2 A12385	peroxidorexin [imp
9	350.5	34.2	200	2 S76284	hypothetical prote
10	345.5	33.7	183	2 A63983	2-cys peroxidorexi
11	336	32.8	178	2 S29119	hypothetical prote
12	330.5	32.3	181	2 A16138	2-cys peroxidorexi
13	328.5	32.1	181	2 A1275	2-cys peroxidorexi
14	325.5	31.8	199	2 F83540	probable alkyl hyd
15	325.5	31.8	200	2 A60552	probable peroxidase
16	322	31.4	226	2 S43598	mers homolog R07E5
17	320.5	31.3	199	2 A48513	macrophage 23k str
18	320.5	31.3	200	2 A80388	probable alkyl hyd
19	319.5	31.2	199	2 I52425	probable thioresox
20	318	31.1	257	2 U00064	MEN's protein - mou
21	317.5	31.0	265	2 T09211	bael protein - spi
22	316.5	30.9	198	2 C64715	alkyl hydroperoxid
23	316.5	30.9	242	2 S49173	hypothetical prote
24	315	30.8	198	2 I68897	probable thioresox
25	314.5	30.7	198	2 H71801	probable peroxidase
26	314.5	30.7	199	2 A46711	probable thioresox
27	314.5	30.7	210	2 T06318	thiol-specific ant
28	314	30.7	192	2 T41413	thioresoxin peroxi
29	312	30.5	198	2 A57716	thiol-specific ant

30	311	30.4	576	2 T16005	hypothetical prote
31	304	29.7	195	2 J02258	substrate protein
32	303	29.6	197	2 C84951	alkyl hydroperoxid
33	302.5	29.5	200	2 I51016	proliferation asso
34	302	29.5	200	2 G83204	probable peroxidase
35	300.5	29.5	198	2 B81453	alkyl hydroperoxid
36	294.5	28.8	196	2 A47362	thiol-specific ant
37	294.5	28.8	202	2 C97756	thioresoxin peroxi
38	291.5	28.5	200	2 A71689	29k peripheral mem
39	289.5	28.3	273	2 A43662	alkyl hydroperoxid
40	284.5	27.8	233	2 S67947	antioxidant; Ahpc/
41	282.5	27.6	207	2 E82287	probable thio-spec
42	281	27.4	195	2 G71492	probable alkyl hyd
43	279	27.2	188	2 D71314	hypothetical prote
44	278.5	27.2	196	2 S69732	alkyl hydroperoxid
45	276.5	27.0	187	2 D87610	

ALIGNMENTS

RESULT 1  
S71013  
alkyl hydroperoxidase reductase chain C - Mycobacterium bovis  
C/Species: Mycobacterium bovis  
C/Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 09-Jul-2004  
C/Accession: S71013; S71011  
R/Wilson, T.M.; Collins, D.M.,  
Mol. Microbiol. 19, 1025-1034, 1996  
A/Title: ahpc, A gene involved in isoniazid resistance of the Mycobacterium tuberculosis  
A/Reference number: S71011; MUID:96249696; PMID:8830260  
A/Accession: S71013  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995  
A/Molecule: nucleic acid sequence not shown; translation not shown  
A/Residues: 1-195 <MW>  
A/Cross-references: UNIPROT:Q57348; UNIPARC:UP1000002FC7F; EMBL:U24083; NID:g1002370; PII  
A/Experimental source: ATCC 35723  
A/Accession: S71013  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995  
A/Molecule type: DNA  
A/Residues: 1-195 <MW>  
A/Cross-references: UNIPARC:UP1000002FC7F; EMBL:U24084; NID:g1002373; PIDN:AA838112.1; P  
A/Experimental source: ATCC 35729  
C/Genetics:  
A/Gene: ahpc  
C/Superfamily: alkyl hydroperoxidase reductase C22 protein; alkyl hydroperoxidase C22 prote  
F123-162/Domain: alkyl hydroperoxidase C22 protein homology <C22>  
Query Match 100.0%; Score 1024; DB 2; Length 195;  
Best Local Similarity 100.0%; Pred. No. 1e-86;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MPLTTGDPFPAYQLTALIGDLSKVDKOPGDTFTTTSDEHPGKRRVVFPMKDTFTFV 60  
DB 1 MPLTTGDPFPAYQLTALIGDLSKVDKOPGDTFTTTSDEHPGKRRVVFPMKDTFTFV 60  
QY 61 CPPEIAAFSKUNEFERDNOIIGVSTDSSEFAHQRWAQHNDDKTLTFPMLSDIKRELISQ 120  
DB 61 CPPEIAAFSKUNEFERDNOIIGVSTDSSEFAHQRWAQHNDDKTLTFPMLSDIKRELISQ 120  
QY 121 AAGVLANADGVADRTFTVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQMRK 180  
DB 121 AAGVLANADGVADRTFTVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQMRK 180  
QY 181 GDPPTLDAGELLKASA 195  
DB 181 GDPPTLDAGELLKASA 195  
RESULT 2  
B70679  
alkyl hydroperoxidase reductase - Mycobacterium tuberculosis (strain H37Rv)  
N/Alternate names: ahpc protein

C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
C;Accession: B70679; S70169  
R;Cole, S.T.; Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Broese, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Raideanu, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544 1998  
A;Authors: Squares, R.; Suleiron, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; PMID:98295987; PMID:9634230  
A;Accession: B70679  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-195 <COL>  
A;Cross-references: UNIPROT:Q57348; UNIPARC:UPI000002FC7F; GB:Z81451; GB:AL123456; NID:9  
R;Experimental source: strain H37Rv  
R;Deretic, V.; Philipp, W.; DhandaVuthapani, S.; Mudd, M.H.; Curcio, R.; Garbe, T.; Heyt  
Mol. Microbiol. 17, 889-900, 1995  
A;Title: Mycobacterium tuberculosis is a natural mutant with an inactivated oxidative-st  
A;Reference number: S70169; PMID:96123431; PMID:8596438  
A;Accession: S70169  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-195 <DER>  
A;Cross-references: UNIPARC:UPI000002FC7F; EMBL:U16243; NID:g1172077; PIDN:AA43585.1; F  
C;Genetic8  
A;Gene: ahpC  
C;Superfamily: alkyl hydroperoxidase reductase C22 protein; alkyl hydroperoxidase c22 prot  
F;23-162/Domain: alkyl hydroperoxidase c22 protein homology <C22>  
  
Query Match 100.0%; Score 1024; DB 2; Length 195;  
Best Local Similarity 100.0%; Pred. No. 1e-86;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 MPLLITGDQFPAYQLTALIGDLSKYDAKQPGDYFTTTSDEHFGKRVVFWPKDFTFV 60  
Db 1 MPLLITGDQFPAYQLTALIGDLSKYDAKQPGDYFTTTSDEHFGKRVVFWPKDFTFV 60  
Qy 61 CPTEIAAFSKLNDSEFDRDAQIIIGVSIIDSEFAHFORAQNHLKTLPPMLSDIKREL 120  
Db 61 CPTEIAAFSKLNDSEFDRDAQIIIGVSIIDSEFAHFORAQNHLKTLPPMLSDIKREL 120  
Qy 121 AAGVLNADGVADRVTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDCLCACNWRK 180  
Db 121 AAGVLNADGVADRVTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDCLCACNWRK 180  
Qy 181 GDPTLDAGELLKASA 195  
Db 181 GDPTLDAGELLKASA 195  
  
RESULT 3  
A43858  
alkyl hydroperoxidase C (EC 1.6.4.-) - Mycobacterium avium  
N;Alternate names: Avt-3 antigen  
C;Species: Mycobacterium avium  
C;Date: 10-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C;Accession: A43858; T09655  
R;Yamauchi, R.; Matsumoto, K.; Yamazaki, A.; Takahashi, M.; Fukasawa, Y.; Wada, M.; Abe, C  
Infect. Immun. 60, 1210-1216, 1992  
A;Title: Cloning and expression of the gene for the Avt-3 antigen of Mycobacterium avium  
A;Reference number: A43858; PMID:1371765  
A;Accession: A43858  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-195 <YAM>  
A;Cross-references: UNIPROT:Q57413; UNIPARC:UPI000002FC83; GB:U18263; NID:g1040852; PIDN  
A;Note: sequence extracted from NCBI backbone (NCBI:95445, NCBI:95446)  
R;Sherman, D.R.; Sabo, P.J.; Hickey, M.J.; Arain, T.M.; Mahatra, G.G.; Yuan, Y.; Barry,  
Proc. Natl. Acad. Sci. U.S.A. 92, 6625-6629, 1995  
A;Title: Disparate responses to oxidative stress in saprophytic and pathogenic mycobacte  
A;Reference number: Z16801; PMID:95327698; PMID:7604044  
A;Accession: T09655

A;Status: preliminary; translated from GB/EMBL/DBS  
A;Molecule type: DNA  
A;Residues: 1-195 <SHE>  
A;Cross-references: UNIPARC:UPI000002FC83; EMBL:U18263; NID:g1040852; PIDN:AAA79917.1; P  
C;Genetics:  
A;Gene: ahpC  
C;Superfamily: alkyl hydroperoxidase reductase C22 protein; alkyl hydroperoxidase c22 prote  
C;Keywords: oxidoreductase  
F;23-162/Domain: alkyl hydroperoxidase c22 protein homology <C22>  
  
Query Match 91.3%; Score 935; DB 2; Length 195;  
Best Local Similarity 90.3%; Pred. No. 1.6e-78;  
Matches 176; Conservative 6; Mismatches 13; Indels 0; Gaps 0;  
  
Qy 1 MPLLITGDQFPAYQLTALIGDLSKYDAKQPGDYFTTTSDEHFGKRVVFWPKDFTFV 60  
Db 1 MPLLITGDQFPAYQLTALIGDLSKYDAKQPGDYFTTTSDEHFGKRVVFWPKDFTFV 60  
Qy 61 CPTEIAAFSKLNDSEFDRDAQIIIGVSIIDSEFAHFORAQNHLKTLPPMLSDIKREL 120  
Db 61 CPTEIAAFSKLNDSEFDRDAQIIIGVSIIDSEFAHFORAQNHLKTLPPMLSDIKREL 120  
Qy 121 AAGVLNADGVADRVTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDCLCACNWRK 180  
Db 121 AAGVLNADGVADRVTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDCLCACNWRK 180  
Qy 181 GDPTLDAGELLKASA 195  
Db 181 GDPTLDAGELLKASA 195  
  
RESULT 4  
E87164  
alkyl hydroperoxidase reductase [imported] - Mycobacterium leprae  
C;Species: Mycobacterium leprae  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
C;Accession: E87164  
R;Cole, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor  
R.; Davies, R.M.; Devlin, K.; Dutuoy, S.; Feltwell, T.; Frazer, A.; Hamlin, N.; Holroyd,  
Nature 409, 1007-1011, 2001  
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq  
A;Title: Massive gene decay in the leprosy bacillus.  
A;Reference number: A86909; PMID:21128732; PMID:11234002  
A;Accession: E87164  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-195 <STO>  
A;Cross-references: UNIPROT:Q9CBF5; UNIPARC:UPI00000662C; GB:AL450380; NID:g13093656; P  
C;Genetics:  
A;Gene: ahpC  
C;Superfamily: alkyl hydroperoxidase reductase C22 protein; alkyl hydroperoxidase c22 prote  
  
Query Match 88.3%; Score 904; DB 2; Length 195;  
Best Local Similarity 86.7%; Pred. No. 1.1e-75;  
Matches 169; Conservative 12; Mismatches 14; Indels 0; Gaps 0;  
  
Qy 1 MPLLITGDQFPAYQLTALIGDLSKYDAKQPGDYFTTTSDEHFGKRVVFWPKDFTFV 60  
Db 1 MPLLITGDQFPAYQLTALIGDLSKYDAKQPGDYFTTTSDEHFGKRVVFWPKDFTFV 60  
Qy 61 CPTEIAAFSKLNDSEFDRDAQIIIGVSIIDSEFAHFORAQNHLKTLPPMLSDIKREL 120  
Db 61 CPTEIAAFSKLNDSEFDRDAQIIIGVSIIDSEFAHFORAQNHLKTLPPMLSDIKREL 120  
Qy 121 AAGVLNADGVADRVTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDCLCACNWRK 180  
Db 121 AAGVLNADGVADRVTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDCLCACNWRK 180  
Qy 181 GDPTLDAGELLKASA 195  
Db 181 GDPTLDAGELLKASA 195



A>Note: Nostrac sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
 C/Accession: A12385  
 R/Kaneke, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasaemoto, S.; Matanabe, A.; Iriyuchi, Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
 DNA Res. 8, 205-213, 2001  
 A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
 A/Reference number: AB1807; MUID:21595285; PMID:11759840  
 A/Accession: A12385  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-203 <KUR>  
 A/Cross-references: UNIPROT:Q8VNC5; UNIPARC:UPI00000CEBD4; GB:BA000019; PIDN:BA076340.1;  
 A/Experimental source: strain PCC 7120  
 C/Genetics:  
 C/Superfamily: alkyl hydroperoxide reductase C22 protein; alkyl hydroperoxidase c22 prot

Query Match 34.3%; Score 351.5; DB 2; Length 203;  
 Best Local Similarity 40.9%; Pred. No. 6.7e-25;  
 Matches 79; Conservative 34; Mismatches 55; Indels 25; Gaps 5;

Qy 4 LTIDGPPAYQLTALIGDLSKVDKAKQPGDYFTTITSDHFGKRVVFFPKDFTVCPT 63  
 Db 11 LRVGQAPDPTATAVVQGE-----FKTKLSDRGKGVVLFYFDLFTVCPT 58

Qy 64 EIAAFSKLNDPEFRDAQILGVSIDSEFAHFQWRAQNDLKT-----LPPMLSDIKREL 118  
 Db 59 EITAFSDRYEERFKKLTIELGVSVDSEFSLAW--IQDRKSGGVGLNLYPLVSDIKKEV 116

Qy 119 SQAGVUN-ADGVADRTFTYDPNNHIOFVSATAGSVGRNDEVLRVLDALQ-----SDE 172  
 Db 117 SDATNVDPAAGIALRGFIIDKGIQHTATINNLAFGRSVDELTKLQALQVQSHPDE 176

Qy 173 LCACNMRKGPDTL 185  
 Db 177 VCPAGMGPGEKTM 189

RESULT 9  
 S76284  
 hypothetical protein - Synechocystis sp. (strain PCC 6803)  
 C/Species: Synechocystis sp.  
 A/Variety: PCC 6803  
 C/Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
 C/Accession: S76284  
 R/Kaneke, T.; Sato, S.; Korani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
 o, K.; Okumura, S.; Shimpou, S.; Takeuchi, C.; Wada, T.; Matanabe, A.; Yamada, M.; Yasuda  
 DNA Res. 3, 109-116, 1996  
 A>Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
 A/Reference number: S74322; MUID:97061201; PMID:8905231  
 A/Accession: S76284  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-200 <KAN>  
 A/Cross-references: UNIPROT:Q55624; UNIPARC:UPI0000139C78; EMBL:D64000; GB:AB001339; NID  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C/Superfamily: alkyl hydroperoxide reductase C22 protein; alkyl hydroperoxidase c22 prot  
 F;14-155/domain: alkyl hydroperoxidase c22 protein homology <C22>

Query Match 34.2%; Score 350.5; DB 2; Length 200;  
 Best Local Similarity 38.5%; Pred. No. 8.2e-25;  
 Matches 74; Conservative 40; Mismatches 57; Indels 21; Gaps 5;

Qy 3 LVTIGDPPAYQLTALIGDLSKVDKAKQPGDYFTTITSDHFGKRVVFFPKDFTVCPT 62  
 Db 4 VLRGQAPDPTATAVVQGE-----FKTKLSDRGKGVVLFYFDLFTVCPT 51

Qy 63 TEIAAFSKLNDPEFRDAQILGVSIDSEFAHFQW---RAQNDLKTLPFPMLSDIKREL 119  
 Db 52 TEIAAFSDRYEERFKKLTIELGVSVDSEFSLAWIQERKMGIGGININPLVSDIKKEIS 111

Qy 120 QAGVUNAD-GVADRTFTYDPNNHIOFVSATAGSVGRNDEVLRVLDALQ-----SDEL 173  
 Db 112 QAVNVLEPPDAGIALRGFIIDREGIQLQYATVNNLSFGRSVDELTKLQALRHVQSHENEV 171

Qy 174 CACNMRKGPDTL 185  
 Db 172 CPVDMDGSDKTM 183

RESULT 10  
 A83983  
 2-cys peroxiredoxin BH265 [imported] - Bacillus halodurans (strain C-125)  
 C/Species: Bacillus halodurans  
 C/Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004  
 C/Accession: A83983  
 R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiran  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A/Reference number: A83650; MUID:20512582; PMID:11058132  
 A/Accession: A83983  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-183 <STO>  
 A/Cross-references: UNIPROT:Q9K912; UNIPARC:UPI00000C3F86; GB:AP001516; GB:BA000004; NID:  
 A/Experimental source: strain C-125  
 C/Genetics:  
 A/Gene: BH265  
 C/Superfamily: alkyl hydroperoxide reductase C22 protein; alkyl hydroperoxidase c22 prote

Query Match 33.7%; Score 345.5; DB 2; Length 183;  
 Best Local Similarity 39.2%; Pred. No. 2.1e-24;  
 Matches 73; Conservative 30; Mismatches 66; Indels 17; Gaps 4;

Qy 6 IDGPPAYQLTALIGDLSKVDKAKQPGDYFTTITSDH--PGKRVVFFPKDFTVCPT 63  
 Db 7 VAKQAPRFEM-----DAVMENKRGKVSLENNKNDKMTVLFYFPMDFVCPT 55

Qy 64 EIAAFSKLNDPEFRDAQILGVSIDSEFAHFQW---RAQNDLKTLPFPMLSDIKREL 120  
 Db 56 EITSLSDRYEFEDLDAEVLGVSTDTHTKAWINTSRDNGIDGLKPYLAADTNHSEVS 115

Qy 121 AAGVUN-ADGVADRTFTYDPNNHIOFVSATAGSVGRNDEVLRVLDALQSDLCACNMR 179  
 Db 116 EYGVLLBEBGIALRGFIISPEGLMYSVNNHNNIGRDVETLRVLDALQGLCPANMK 175

Qy 180 KGDPDTL 185  
 Db 176 PGQETL 181

RESULT 11  
 S29119  
 hypothetical protein 3 - Clostridium pasteurianum  
 C/Species: Clostridium pasteurianum  
 C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004  
 C/Accession: S29119  
 R/Mathieu, I.; Meyer, J.; Moullis, J.M.  
 Biochem. J. 285, 255-262, 1992  
 A>Title: Cloning, sequencing and expression in Escherichia coli of the rubredoxin gene fr  
 A/Reference number: S29117; MUID:92344580; PMID:1637309  
 A/Accession: S29119  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-178 <MAT>  
 A/Cross-references: UNIPROT:P23161; UNIPARC:UPI000002FC06; EMBL:M60116; NID:G144905; PIDN  
 C/Superfamily: alkyl hydroperoxide reductase C22 protein; alkyl hydroperoxidase c22 prote  
 F;12-155/domain: alkyl hydroperoxidase c22 protein homology <C22>

Query Match 32.8%; Score 336; DB 2; Length 178;  
 Best Local Similarity 38.6%; Pred. No. 1.5e-23;  
 Matches 71; Conservative 34; Mismatches 65; Indels 14; Gaps 3;

Qy 6 IDGPPAYQLTALIGDLSKVDKAKQPGDYFTTITSDHFGKRVVFFPKDFTVCPT 61

```
Db      5 VGRKAPFEEMKAV-----KGDGKGFTEVLGKGMVLWFFPLDFTVCPEI 54
Qy      66 AAFSKLNDPEFDRDAQILGVISIDSEFAHFQWR---HNDKTLPPFMLSIDIKRELSDA 122
      55 TGFESKRAEEFDDLKAEILLVSCDSQYSHETINDIKOGGIGKLNFPPLASXTTEVSTKY 114
Qy      123 GV-LNADGVADRVTFTVDPNNEIQFVSATAGSVGRNDEVLRVLDALQSDDELCAQNRKG 181
      115 GIGIEEGISLRGFIIDPEGI VRYSVVHDLNVRSDVETLRVLFKAPGTGMCALDMHEG 174
Db      182 DPTL 185
      175 DDNL 178

RESULT 12
AD1638
2-cys peroxiredoxin homolog l1n1645 [imported] - Listeria innocua (strain Clp11262)
C/Species: Listeria innocua
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C/Accession: AD1638
R/Glaeser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurge, O.; Entian, K.D.; Feihl, H.
Science 294, 849-852, 2001
A/Authors: Krefit, J.; Simoes, N.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Me
ok, C.; Schluter, T.; Simeos, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A./Title: Comparative genomics of Listeria species.
A/Reference number: AB1077; MUID:21537279; PMID:11679669
A/Accession: AD1638
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-181 <GLA>
A/Cross-references: UNIPROT:Q92BA5; UNIPARC:UPI000000CC601; GB:AL592022; PIDN:CAC96876.1;
A/Experimental source: strain Clp11262
C/Genetics:
A/Gene: l1n1645
C/Superfamily: alkyl hydroperoxide reductase C22 protein; alkyl hydroperoxidase c22 pro
Query Match      32.3%; Score 330.5; DB 2; Length 181;
Best Local Similarity 37.1%; Pred. No. 4.9e-23;
Matches 69; Conservative 38; Mismatches 62; Indels 17; Gaps 4;

Qy      6 IGDQFPAYQLTALIGDLSKYDAKQPGDYFTTTSDEH--PGKRVVFEMPKDFTVCPT 63
      6 VGTQAPFEEMKAVM-----PNQTFGKVSLEKNIEDDKTWLTFEYPMDFTVCP 54
Db      64 EIAAFSKLNDPEFDRDAQILGVISIDSEFAHFQWR---AQHNDKTLPPFMLSIDIKRELSDQ 120
      55 EIVAIASRSDEFDLAKRIITGASTDTTSHLAWNTPTIKGGIGKLNFPPLADTNHGVAS 114
Qy      121 AAGVL-NADGVADRVTFTVDPNNEIQFVSATAGSVGRNDEVLRVLDALQSDDELCAQNR 179
      115 DYGVLIEEGEVALRGLFTINPKGEIQEVVHNNIGREDEVLRVQLQGTGCLPIWQ 174
Db      180 KGDP 185
      175 PGEKTI 180

RESULT 13
AD1275
2-cys peroxiredoxin homolog lmo1604 [imported] - Listeria monocytogenes (strain EGD-e)
C/Species: Listeria monocytogenes
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C/Accession: AD1275
R/Glaeser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurge, O.; Entian, K.D.; Feihl, H.
Science 294, 849-852, 2001
A/Authors: Krefit, J.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simeos, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
```

```
A/Title: Comparative genomics of Listeria species.
A/Reference number: AB1077; MUID:21537279; PMID:11679669
A/Accession: AD1275
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-181 <GLA>
A/Cross-references: UNIPROT:Q8Y6S9; UNIPARC:UPI000000CF186; GB:NC_003210; PIDN:CAC99682.1;
A/Experimental source: strain EGD-e
C/Genetics:
A/Gene: lmo1604
C/Superfamily: alkyl hydroperoxide reductase C22 protein; alkyl hydroperoxidase c22 pro
Query Match      32.1%; Score 328.5; DB 2; Length 181;
Best Local Similarity 36.6%; Pred. No. 7.5e-23;
Matches 68; Conservative 39; Mismatches 62; Indels 17; Gaps 4;

Qy      6 IGDQFPAYQLTALIGDLSKYDAKQPGDYFTTTSDEH--PGKRVVFEMPKDFTVCPT 63
      6 VGTQAPFEEMKAVM-----PNQTFGKVSLEKNIEDDKTWLTFEYPMDFTVCP 54
Db      64 EIAAFSKLNDPEFDRDAQILGVISIDSEFAHFQWR---AQHNDKTLPPFMLSIDIKRELSDQ 120
      55 EIVAIASRSDEFDLAKRIITGASTDTTSHLAWNTPTIKGGIGKLNFPPLADTNHGVAS 114
Qy      121 AAGVL-NADGVADRVTFTVDPNNEIQFVSATAGSVGRNDEVLRVLDALQSDDELCAQNR 179
      115 DYGVLIEEGEVALRGLFTINPKGEIQEVVHNNIGREDEVLRVQLQGTGCLPIWQ 174
Db      180 KGDP 185
      175 PGEKTI 180

RESULT 14
F83540
probable alkyl hydroperoxide reductase PA0848 [imported] - Pseudomonas aeruginosa (strain
C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C/Accession: F83540
R/Stover, C.K.; Pham, X.O.; Ervin, A.L.; Mitsuuchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbys, K.; Lim,
Lozy, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathoc
A/Reference number: AB2950; MUID:20437337; PMID:10984043
A/Accession: F83540
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-199 <STO>
A/Cross-references: UNIPROT:Q91593; UNIPARC:UPI000000C5187; GB:AE004519; GB:AE004091; NID:
A/Experimental source: strain PA01
C/Genetics:
A/Gene: PA0848
C/Superfamily: alkyl hydroperoxide reductase C22 protein; alkyl hydroperoxidase c22 pro
Query Match      31.8%; Score 325.5; DB 2; Length 199;
Best Local Similarity 37.1%; Pred. No. 1.6e-22;
Matches 73; Conservative 39; Mismatches 62; Indels 23; Gaps 5;

Qy      6 IGDQFPAYQLTALIGDLSKYDAKQPGDYFTTTSDEHPGKRVVFEMPKDFTVCPT 65
      5 VNRQAPDFTAAAVL-GDGSIVDAFQLSSL-----RGKYVVLFPWPIDFTVCPSEI 54
Db      66 AAFSKLNDPEFDRDAQILGVISIDSEFAHFQWR---QHNDKTLPPFMLSIDIKRELSDA 122
      55 IAHNNRDKREIEGVEVGVISIDQFTTHAMRSTPVKGGIGAEVFPVADVGHETIRAY 114
Qy      123 GV-LNADGVADRVTFTVDPNNEIQFVSATAGSVGRNDEVLRVLDALQ---SDELCAQNR 178
      115 GIEHEDVALRASGLIRAGVGHQVNNPLPLGREDVEMRLVLEALQFTEHGHVCPAGW 174
Db      179 RKGDPTLAGELLKASA 195
      |||
```

Db 175 RKQ-----KGMKSA 185

## RESULT 15

AB0552

probable peroxidase [imported] - *Salmonella enterica* subsp. *enterica* serovar Typh1 (stre

C:Species: *Salmonella enterica* subsp. *enterica* serovar Typh1

A:Note: this species has also been called *Salmonella typhi*

C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C:Accession: AB0552

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AB0552

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-200 <PAR>

A:Cross-references: UNIPARC:UP1000005A247; GB:AL513382; PIDN:CAD08858.1; PID:G16501671;

C:Genetics:

A:Gene: STY0440

C:Superfamily: alkyl hydroperoxide reductase C22 protein; alkyl hydroperoxidase c22 prot

Query Match 31.8%; Score 325.5; DB 2; Length 200;

Best Local Similarity 35.6%; Pred. No. 1.6e-22;

Matches 69; Conservative 39; Mismatches 67; Indels 19; Gaps 5;

QY 3 LRTIGDQFPAYQLTALIGDLSKVDAKQPGDYFTTTSDEHP-GKMRVFPWPKDFTFVC 61

Db 1 MVLVTRQAPDPTAAAVLGS-----GEIVDKFNFKQHTNGKTTVLFFPMDFTFVC 50

QY 62 PTEIAFSLKLNDEFEDRDAQILGVISIDSEFAHFQMR--AQHNDLKTLPFMLSIDIKREL 118

Db 51 PSELIAFDKRYEERQKRGVEVGVSPSEFVHNAMKNTLPVDKSGIGGVKXAMADVGREI 110

QY 119 SQAGVLNAD-GVADRVTFIVDPNNEIQVSATAGVGRNVDEVLRVLAIQSDE---L 173

Db 111 OKAYGIEHPDEGVLRGSLIDANGIVRHQGVVNDLPLGRNIDEMLRVVDALQFHEHGDV 170

QY 174 CACNWRKGDPTLDA 187

Db 171 CPAQWEKKEGEMNA 184

Search completed: March 23, 2006, 05:26:32  
Job time : 7.75813 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 04:56:43 ; Search time 49.8737 Seconds  
(without alignments)

2758.529 Million cell updates/sec

Title: US-10-617-038-26

Perfect score: 1024

Sequence: 1 MPLLTIGDPPAYQLTALIG.....CNMRKDDPTLDAGELKASA 195

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1024	100.0	195	2	Q79CVO MYCBO
2	1024	100.0	195	2	Q7BHK8 MYCTU
3	943	92.1	195	2	087323 MYCNR
4	935	91.3	195	2	057413 MYCAV
5	934	91.2	195	2	Q732L3 MYCPA
6	920	89.8	195	2	Q9AGF6 MYCPA
7	913	89.2	195	2	05YTS4 NOCPA
8	904	88.3	195	2	09CBF5 MYCLE
9	878	85.7	195	2	057529 MYCSM
10	693	67.7	198	2	Q6NGT3 CORDI
11	685	66.9	198	2	Q460Z5 CORDI
12	619	60.4	194	2	Q9XSV0 STRVD
13	593	57.9	184	2	Q9FBP5 STRCO
14	592	57.8	184	2	Q9RNT2 STRCO
15	589	57.6	184	2	0821C6 STRAV
16	589	57.5	178	2	Q4NTM9 SDELT
17	569.5	55.6	179	2	083BM6 COXBU
18	566.5	55.3	179	2	05X2T8 LEGPL
19	565.5	55.2	179	2	05WU96 LEGPL
20	564.5	55.1	212	2	05ZTJ5 LEGPH
21	455.5	44.5	180	2	05NL04 ZYMMO
22	453	44.2	184	2	Q578J3 BRUBA
23	453	44.2	184	2	08FVW3 BRUSU
24	444	43.4	184	2	08YCF3 BRUME
25	444	43.4	184	2	Q9ANK9 BRALJA
26	429.5	41.9	230	2	08XYW0 RALSO
27	423.5	41.4	182	2	062J18 BURMA
28	423.5	41.4	182	2	063T73 BURPS
29	410.5	40.1	182	2	Q7VTT5 BORPE
30	410.5	40.1	182	2	Q7W5D3 BORPA
31	410.5	40.1	182	2	Q7WCW8 BORBR

32	404	39.5	181	2	Q5QZB4 IDLIO	Q5qz24 idiomarina
33	389	38.0	179	2	Q67QY2 SYMTH	Q67qy2 symbiobacte
34	374	36.5	180	2	Q74ER7 GEOSTL	Q74er7 geobacter s
35	371	36.2	194	2	Q7VIX9 PROMP	Q7vix9 prochloroco
36	361.5	35.3	180	2	034564 BACSU	034564 bacillus su
37	360.5	35.2	180	2	065K82 BACAD	065k82 bacillus li
38	360	35.2	189	1	YCF42 FORPU	Ycf42 porphyra pu
39	358.5	35.0	200	2	Q7V7J1 PROMM	Q7v7j1 prochloroco
40	357	34.9	197	2	Q7VEM4 PROMA	Q7vem4 prochloroco
41	353.5	34.5	200	2	Q7VEX3 SYNXP	Q7vex3 synechococc
42	352	34.4	197	2	08DIX7 SYNEL	08dix7 synechococc
43	351.5	34.3	203	2	Q8YNC5 TANAS	Q8ync5 anabaena sp
44	351	34.3	199	2	Q7KGY6 LEICH	Q7kgy6 leishmania
45	351	34.3	199	2	Q95NFS LEIIN	Q95nfs leishmania

## ALIGNMENTS

Q79CVO MYCBO	PRELIMINARY:	PRT: 195 AA.
ID	Q79CVO MYCBO	Q79CVO MYCBO
AC	Q79CVO_050557; Q50558; Q50572; Q50573; Q50574; Q57348;	
DT	05-JUL-2004 (TrEMBLrel. 27, Created)	
DT	05-JUL-2004 (TrEMBLrel. 27, Last sequence update)	
DT	13-SEP-2005 (TrEMBLrel. 31, Last annotation update)	
DE	Alkyl hydroperoxide reductase C peptide (ALKYL HYDROPEROXIDE REDUCTASE	
DE	C PROTEIN AHPC (ALKYL HYDROPEROXIDASE C) (EC 1.-.-.-.)	
GN	Name=ahpc; OrderedLocusNames=Mb2454;	
OS	Mycobacterium bovis.	
OC	Bacteria; Actinobacteria; Actinomycetales;	
OC	Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;	
OC	Mycobacterium tuberculosis complex.	
OX	NCBI_TaxID=1765;	
RN	[1]	
RP	NUCLEOTIDE SEQUENCE.	
RC	STRAIN=ATCC 35723, and ATCC 35729;	
RA	MEDLINE=9624966; PubMed=8030260;	
RT	Wilson T.M., Collins D.M.;	
RT	"ahpc, a gene involved in isoniazid resistance of the Mycobacterium	
RT	tuberculosis complex.";	
RL	Mol. Microbiol. 19:11025-1034(1996).	
RN	[2]	
RP	NUCLEOTIDE SEQUENCE.	
RC	STRAIN=AF2122/97;	
RA	MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;	
RA	Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,	
RA	Pryor M., Dutfoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,	
RA	Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,	
RA	Parikhil J., Barrett B.G., Cole S.T., Gordon S.V., Hewinson R.G.;	
RT	"The complete genome sequence of Mycobacterium bovis.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).	
DR	EMBL; U24083; AAB60203.1; -; Genomic DNA.	
DR	EMBL; U24084; AAB38112.1; -; Genomic DNA.	
DR	EMBL; BX248342; CAD97315.1; -; Genomic DNA.	
DR	PIR; B70679; B70679.	
DR	PIR; S71013; S71013.	
DR	SMR; Q79CVO; 2-179.	
DR	GO; GO:0016491; F:oxidoreductase activity; IEA.	
DR	GO; GO:0004601; F:peroxidase activity; IEA.	
DR	InterPro; IPR000866; Ahpc-TSA.	
DR	InterPro; IPR012336; Thiooxidin-like.	
DR	Pfam; PF00578; Ahpc-TSA; 1.	
KM	Complete proteome; Oxidoreductase; Peroxidase.	
SQ	SEQUENCE 195 AA; 21566 MW; 011C101F07C7095 CRC64;	
Q7	Query Match	100.0%; Score 1024; DB 2; Length 195;
Db	Best Local Similarity	100.0%; Pred. No. 8.7e-82;
	Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
	1 MPLLTIGDPPAYQLTALIGDLSKVDKAPGDIYTTTSDHPGKRVVFPWPKDFTFV 60	



QY 61 CPTBIAAFSKLNDEFEDRDAQIIIGVSIIDSEFAHFQWRAQHNDKTLPPMLSDIKRELISQ 120  
 DB 61 CPTBIAAFSKLNDEFEDRDAQIIIGVSIIDSEFAHFQWRAQHNDKTLPPMLSDIKRELISQ 120  
 QY 121 AAGVLNADGVADRVTVIVDPNNELQFVSATAGSGVRNDEVLRVLDALQSDLCACNWRK 180  
 DB 121 AAGVLNADGVADRVTVIVDPNNELQFVSATAGSGVRNDEVLRVLDALQSDLCACNWRK 180  
 QY 181 GDPTLDAGELLKASA 195  
 DB 181 GDPTLDAGELLKASA 195

RESULT 2  
 ID 07BHK8 MYCTU PRELIMINARY; PRT; 195 AA.  
 AC 07BHK8 079FE2: 07D758:  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 31, Last annotation update)  
 DE Alkyl hydroperoxidase C (ALKYL HYDROPEROXIDE REDUCTASE C PROTEIN AHPC (ALKYL HYDROPEROXIDASE C) (EC 1.-.-.-)) (Ahpc).  
 GN Name=ahpc; OrderedLocNames=WT5503, Rv2428;  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;  
 OC Mycobacterium tuberculosis complex.  
 OK NCBI\_Taxid=1773;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Orru G., Iona E., Memmi G., Oggioni M.R., Factorini L., Orefici G., Pozzi G.;  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230; DOI=10.1038/31159;  
 RA Cole S.T., Broesch R., Parhill J., Garnier T., Churcher C.M., Harris D.E., Gordon S.V., Eigmeier K., Gae S., Barry C.E. III, Tekala F., Badcock K., Baaham D., Brown D., Chillingworth T., Connor R., Davies R.M., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S., Hornsby T., Jagers K., Krogan A., McLean J., Moule S., Murphy L.D., Oliver S., Osborne K., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sultson J.E., Taylor S., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence";  
 RT Nature 393:537-544(1998).  
 RL Nature 393:537-544(1998).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RX MEDLINE=22206494; PubMed=12218036;  
 RX DOI=10.1128/JB.184.19.5479-5490.2002;  
 RA Fietechmann R.D., Alland D., Eilen J.A., Carpenter L., White O., Peterson J.D., Deboy R.T., Dodson R.J., Gwyn L.W., Haft D.H., Hickey E.K., Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Slatberg S.L., Delcher A., Uterback T.R., Waldman J.F., Khouri H.M., Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
 RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains";  
 RT J. Bacteriol. 184:5479-5490(2002).  
 RL J. Bacteriol. 184:5479-5490(2002).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=96123431; PubMed=8596438;  
 RX Dereic V., Philipp W., Dhanayathapani S., Mudd M.H., Curcio R., Garbe T., Heym B., Via L.E., Cole S.T.;  
 RA "Mycobacterium tuberculosis is a natural mutant with an inactivated oxidative-stress regulatory gene: implications for sensitivity to isoniazid";  
 RT Mol. Microbiol. 17:889-900(1995).

RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=95327698; PubMed=7604044;  
 RA Sherman D.R., Sabo P.J., Hickey M.J., Arain T.M., Mahatras G.G., Yuan Y., Barry C.E. III, Stover C.K.;  
 RT "Disparate responses to oxidative stress in saprophytic and pathogenic Mycobacteria";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:6625-6629(1995).  
 DR EMBL: AF313459; AAC34172.1; -; Genomic\_DNA.  
 DR EMBL: BX842579; CAB03768.1; -; Genomic\_DNA.  
 DR EMBL: AF313460; AAC34173.1; -; Genomic\_DNA.  
 DR EMBL: AF000516; AA46800.1; -; Genomic\_DNA.  
 DR EMBL: AF313461; AAC34174.1; -; Genomic\_DNA.  
 DR EMBL: AF313463; AAC34176.1; -; Genomic\_DNA.  
 DR EMBL: U18264; AA479919.1; -; Genomic\_DNA.  
 DR EMBL: U16243; AAC43585.1; -; Genomic\_DNA.  
 DR EMBL: AF313462; AAC34175.1; -; Genomic\_DNA.  
 DR PDB: 2BMX; X-ray; A/B/C=1-195.  
 DR SMR; 07BHK8; 2-179.  
 DR TIGR; MT503; -;  
 DR Tuberculist; Rv2428; -;  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0004601; F:peroxidase activity; IEA.  
 DR InterPro: IPR000865; Ahpc-TSA.  
 DR InterPro: IPR012335; Thioresoxin-like.  
 DR InterPro: IPR012335; Thioresoxin-like.  
 DR Pfam; PF00578; Ahpc-TSA; 1.  
 KW Complete proteome; Oxidoreductase; Peroxidase.  
 SQ SEQUENCE 195 AA; 21566 MW; 011C1014F07C7095 CRC64;

Query Match 100.0%; Score 1024; DB 2; Length 195;  
 Best Local Similarity 100.0%; Pred. No. 8; 7e-82;  
 Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFLTTIGDOPPAQQLALIGGDSKVDKOPGPFYTTISDEHGRVVFYFPKQFTFY 60  
 DB 1 MFLTTIGDOPPAQQLALIGGDSKVDKOPGPFYTTISDEHGRVVFYFPKQFTFY 60  
 QY 61 CPTBIAAFSKLNDEFEDRDAQIIIGVSIIDSEFAHFQWRAQHNDKTLPPMLSDIKRELISQ 120  
 DB 61 CPTBIAAFSKLNDEFEDRDAQIIIGVSIIDSEFAHFQWRAQHNDKTLPPMLSDIKRELISQ 120  
 QY 121 AAGVLNADGVADRVTVIVDPNNELQFVSATAGSGVRNDEVLRVLDALQSDLCACNWRK 180  
 DB 121 AAGVLNADGVADRVTVIVDPNNELQFVSATAGSGVRNDEVLRVLDALQSDLCACNWRK 180  
 QY 181 GDPTLDAGELLKASA 195  
 DB 181 GDPTLDAGELLKASA 195

RESULT 3  
 ID 087323 MYCWR PRELIMINARY; PRT; 195 AA.  
 AC 087323;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DE Alkylhydroperoxide reductase.  
 GN Name=ahpc;  
 OS Mycobacterium marinum.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OK NCBI\_Taxid=1781;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=ATCC 15069;  
 RX MEDLINE=98406038; PubMed=9733688;  
 RX Pagan-Ramos E., Song J., McFalone M., Mudd M.H., Dereic V.;  
 RA "Oxidative stress response and characterization of the oxyR-ahpc and furA-katG loci in Mycobacterium marinum";  
 RT J. Bacteriol. 180:4856-4864(1998).  
 RL EMBL: AF034861; AAC61301.1; -; Genomic\_DNA.

DR HSP, Q9TX2, 1E2Y.  
 DR SMK; O87323; 3-179.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0004601; F:peroxidase activity; IEA.  
 DR InterPro; IPR000866; Ahpc-TSA.  
 DR InterPro; IPR012336; Thioredoxin-like.  
 DR InterPro; IPR012335; Thioredoxin\_fold.  
 DR Pfam; PF00578; Ahpc-TSA; 1.  
 SQ SEQUENCE 195 AA; 21440 MW; 4259767C92A9D31 CRC64;

Query Match 92.1%; Score 943; DB 2; Length 195;  
 Best Local Similarity 90.8%; Pred. No. 1.1e-74;  
 Matches 177; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 MPLLITIDQFPAYQLTALIGDLSKVDKOPGDIYFTTTSDEHGKRRVVFPMKDTFTV 60  
 DB 1 MSLLITIQQFPAYSLTLALIGDLSKVDKOPGDIYFTTTSDEHGKRRVVFPMKDTFTV 60  
 QY 61 CPTETIAAFSKLNDFEDRDAQILGVSIIDSEFAHFQWRAGHNDLKTLPFMLSIDIKRELISQ 120  
 DB 61 CPTETIAAFGKLNDFEDRDAQVLGVSIIDSEFVHFQWRAGHNDLKTLPFMLSIDIKRELISQ 120  
 QY 121 AAGVLNADGVADRTFTVDPNNELQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQWRRK 180  
 DB 121 ATGVLANADGVADRTFTVDPNNELQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQWRRK 180  
 QY 181 GDPTLDAGEILKASA 195  
 DB 181 GDPTLDAGEILKASA 195

RESULT 4  
 057413 MYCAV PRELIMINARY; PRT; 195 AA.

AC 057413; 1-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)  
 DE Alkyl hydroperoxidase C (Antigen).  
 GN Name=ahpc; Synonyms=Avl-3;  
 OS Mycobacterium avium.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;  
 OC Mycobacterium avium complex (MAC).  
 CX NCBI\_TaxID=1764;  
 RN NUCLEOTIDE SEQUENCE.  
 RP MEDLINE=95327698; PubMed=7604044;  
 RA Sherman D.R., Sabo P.J., Hickey M.J., Arain T.M., Mahairas G.G.,  
 RA Yuan Y., Barry C.E. III, Stover C.K.;  
 RT "Disparate responses to oxidative stress in saprophytic and pathogenic  
 RT Mycobacteria";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:6625-6629(1995).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RP MEDLINE=92175967; PubMed=1371765;  
 RA Yamaguchi R., Matsuo K., Yamazaki A., Takahashi M., Fukasawa Y.,  
 RA Mada M., Abe C.;  
 RT "Cloning and expression of the gene for the Avl-3 antigen of  
 RT Mycobacterium avium and mapping of its epitopes";  
 RL Infect. Immun. 60:1210-1216(1992).  
 DR EMBL; U1863; AAA79917.1; -; Genomic DNA.  
 DR EMBL; M74232; AAA25357.1; -; Genomic DNA.  
 DR HSP; P32119; 10MV.  
 DR SMK; Q57413; 2-179.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR InterPro; IPR000866; Ahpc-TSA.  
 DR InterPro; IPR012336; Thioredoxin-like.  
 DR InterPro; IPR012335; Thioredoxin\_fold.  
 DR Pfam; PF00578; Ahpc-TSA; 1.  
 KM Peroxidase.

SQ SEQUENCE 195 AA; 21655 MW; 11A60C1849343FD7 CRC64;

Query Match 91.3%; Score 935; DB 2; Length 195;  
 Best Local Similarity 90.3%; Pred. No. 5.7e-74;  
 Matches 176; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 1 MPLLITIDQFPAYQLTALIGDLSKVDKOPGDIYFTTTSDEHGKRRVVFPMKDTFTV 60  
 DB 1 MPLLITIDQFPAYVELTLALIGDLSKVDKOPGDIYFTTTSDEHGKRRVVFPMKDTFTV 60  
 QY 61 CPTETIAAFSKLNDFEDRDAQILGVSIIDSEFAHFQWRAGHNDLKTLPFMLSIDIKRELISQ 120  
 DB 61 CPTETIAFGKLNDFEDRDAQVLGVSIIDSEFVHFQWRAGHNDLKTLPFMLSIDIKRELISQ 120  
 QY 121 AAGVLNADGVADRTFTVDPNNELQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQWRRK 180  
 DB 121 ATGVLANADGVADRTFTVDPNNELQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQWRRK 180  
 QY 181 GDPTLDAGEILKASA 195  
 DB 181 GDPTLDAGEILKASA 195

RESULT 5  
 073ZL3 MYCPA PRELIMINARY; PRT; 195 AA.

AC 073ZL3;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Ahpc.  
 GN Name=ahpc; Ordered locus names=MAP1589c;  
 OS Mycobacterium paratuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;  
 OC Mycobacterium avium complex (MAC).  
 CX NCBI\_TaxID=1770;  
 RN NUCLEOTIDE SEQUENCE.  
 RC STRAIN=K10;  
 RA Li L., Bannantine J., Zhang Q., Amonsin A., Alt D., Kapur V.;  
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE017232; AAS03906.1; -; Genomic DNA.  
 DR SMK; Q73ZL3; 2-179.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0004601; F:peroxidase activity; IEA.  
 DR InterPro; IPR000866; Ahpc-TSA.  
 DR InterPro; IPR012336; Thioredoxin-like.  
 DR Pfam; PF00578; Ahpc-TSA; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 195 AA; 21641 MW; 4A9D9C0BF102E61D CRC64;

Query Match 91.2%; Score 934; DB 2; Length 195;  
 Best Local Similarity 89.7%; Pred. No. 7e-74;  
 Matches 175; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY 1 MPLLITIDQFPAYQLTALIGDLSKVDKOPGDIYFTTTSDEHGKRRVVFPMKDTFTV 60  
 DB 1 MPLLITIDQFPAYVELTLALIGDLSKVDKOPGDIYFTTTSDEHGKRRVVFPMKDTFTV 60  
 QY 61 CPTETIAAFSKLNDFEDRDAQILGVSIIDSEFAHFQWRAGHNDLKTLPFMLSIDIKRELISQ 120  
 DB 61 CPTETIAFGKLNDFEDRDAQVLGVSIIDSEFVHFQWRAGHNDLKTLPFMLSIDIKRELISQ 120  
 QY 121 AAGVLNADGVADRTFTVDPNNELQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQWRRK 180  
 DB 121 ATGVLANADGVADRTFTVDPNNELQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQWRRK 180  
 QY 181 GDPTLDAGEILKASA 195  
 DB 181 GDPTLDAGEILKASA 195

```

RESULT 6
Q9AGF6_MYCPA PRELIMINARY; PRT; 195 AA.
ID Q9AGF6_MYCPA PRELIMINARY; PRT; 195 AA.
AC Q9AGF6;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Alkylhydroperoxidase C.
GN Name=ahpc;
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium avium complex (MAC).
OX NCBI_TaxID=1770;
RN NUCLEOTIDE SEQUENCE.
RP Chang Y.F., Shin K.-S.;
RA Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF34163; AK20392.1; -; Genomic_DNA.
DR HSSP; Q9T2X2; 1E2Y.
DR SMK; Q9AGF6; 2-179.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR InterPro; IPR000866; Ahpc-TSA.
DR InterPro; IPR012336; Thioresoxin-like.
DR InterPro; IPR012335; Thioresoxin_fold.
DR Pfam; PF00578; Ahpc-TSA; 1.
DR Peroxidase.
KW SEQUENCE 195 AA; 21726 MW; 808557FLPASF78B9 CRC64;

Query Match 89.8%; Score 920; DB 2; Length 195;
Best Local Similarity 88.2%; Pred. No. 1.2e-72;
Matches 172; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

QY 1 MLLTTIGDGPAYQLTALIGDLSKVDKQPGDYFTTTSDEHPGKRVVFFMPKDTFV 60
DB 1 MLLTTIGDGPAYQLTALIGDLSKVDKQPGDYFTTTSDEHPGKRVVFFMPKDTFV 60
QY 61 CPTETIAAFSKLNDFFEDRDQIIGVSIIDSEFAHQRQAQNDLKTLPFMLSIDIKREL 120
DB 61 CPTETIAAFSKLNDFFEDRDQIIGVSIIDSEFAHQRQAQNDLKTLPFMLSIDIKREL 120
QY 121 AAGVLNADGVADRTFTVDPNNEIQFVSATAGSVGRVDEVLRTLDALQSDLCACNWK 180
DB 121 AAGVLNADGVADRTFTVDPNNEIQFVSATAGSVGRVDEVLRTLDALQSDLCACNWK 180
QY 181 GDPTLADGELLKASA 195
DB 181 GDPTLADGELLKASA 195
QY 181 GDPTLADGELLKASA 195
DB 181 GDPTLADGELLKASA 195

RESULT 7
Q5YTS4_MYCPA PRELIMINARY; PRT; 195 AA.
ID Q5YTS4_MYCPA PRELIMINARY; PRT; 195 AA.
AC Q5YTS4;
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE Putative alkylhydroperoxide reductase.
GN Name=ahpc; OrderedLocNames=nf37890;
OS Nocardia farcinica.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Corynebacteriaceae; Nocardia.
OX NCBI_TaxID=37329;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=IFM 10152;
RA PubMed=15466710; DOI=10.1073/pnas.0406410101;
RA Ishikawa J., Yamashita A., Mikami Y., Hoshino Y., Kurita H., Hotta K.,
RA Shiba T., Hattori M.;
RT "The complete genomic sequence of Nocardia farcinica IFM 10152."
RL Proc. Natl. Acad. Sci. U.S.A. 101:14925-14930(2004).
DR EMBL; AF006618; BAD58637.1; -; Genomic_DNA.

```

```

DR SMK; Q5YTS4; 3-179.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR InterPro; IPR000866; Ahpc-TSA.
DR InterPro; IPR012336; Thioresoxin-like.
DR Pfam; PF00578; Ahpc-TSA; 1.
KW Complete proteome.
SQ SEQUENCE 195 AA; 21456 MW; 6B3D2F9DC95910C9 CRC64;

Query Match 89.2%; Score 913; DB 2; Length 195;
Best Local Similarity 86.6%; Pred. No. 4.9e-72;
Matches 168; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 1 MLLTTIGDGPAYQLTALIGDLSKVDKQPGDYFTTTSDEHPGKRVVFFMPKDTFV 60
DB 1 MLLTTIGDGPAYQLTALIGDLSKVDKQPGDYFTTTSDEHPGKRVVFFMPKDTFV 60
QY 61 CPTETIAAFSKLNDFFEDRDQIIGVSIIDSEFAHQRQAQNDLKTLPFMLSIDIKREL 120
DB 61 CPTETIAAFSKLNDFFEDRDQIIGVSIIDSEFAHQRQAQNDLKTLPFMLSIDIKREL 120
QY 121 AAGVLNADGVADRTFTVDPNNEIQFVSATAGSVGRVDEVLRTLDALQSDLCACNWK 180
DB 121 AAGVLNADGVADRTFTVDPNNEIQFVSATAGSVGRVDEVLRTLDALQSDLCACNWK 180
QY 181 GDPTLADGELLKASA 194
DB 181 GDPTLADGELLKASA 194

RESULT 8
Q9CBF5_MYCLE PRELIMINARY; PRT; 195 AA.
ID Q9CBF5_MYCLE PRELIMINARY; PRT; 195 AA.
AC Q9CBF5;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Alkyl hydroperoxide reductase.
GN Name=ahpc; OrderedLocNames=ML2042;
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=TN;
RA PubMed=1128732; PubMed=11234002; DOI=10.1038/35059006;
RA Cole S.T., Eigemeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C.M., Harris D.E.,
RA Wungali K.L., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltham A., Hamlin N.,
RA Holroyd S., Hornby T., Jags K., Lacroix C., Maclean J., Moule S.,
RA Murphy L.D., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skellon J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrett B.G.;
RT "Passive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
DR EMBL; AL583924; CAC30997.1; -; Genomic_DNA.
DR PIR; E87164; E87164.
DR HSSP; Q9T2X2; 1E2Y.
DR SMK; Q9CBF5; 3-179.
DR Leproma; ML2042; -;
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR InterPro; IPR000866; Ahpc-TSA.
DR InterPro; IPR012336; Thioresoxin-like.
DR InterPro; IPR012335; Thioresoxin_fold.
DR Pfam; PF00578; Ahpc-TSA; 1.
KW Complete proteome.
SQ SEQUENCE 195 AA; 21497 MW; DB68152F813CEA1 CRC64;

Query Match 88.3%; Score 904; DB 2; Length 195;

```

Best Local Similarity 86.7%; Pred. No. 3e-71;  
Matches 169; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

```

Qy 1 MLLTTIGQPPAYQTLALIGDLSKVDAKOPGDYFTTTSDEHPGKRVVFWMPPKDFTFV 60
Db 1 MSLLSIGQPPAYQTLALIGDLSKVDAKOPGDYFTTTSDEHPGKRVVFWMPPKDFTFI 60
Qy 61 CPTETAAFSKLNDEFERDAQILGVSIDSEFAHQWRAQHNDLKTLPFPMLSDIKRELSQ 120
Db 61 CPTETAAFGKLNDEFERGAQILGVSIDSEFAHQWRAQHNDLKTLPFPMLSDIKRELSA 120
Qy 121 AAGVLNADGVADRTFTYIDPNNEIOFVSATAGSVGRNVDVLRVLDAQSDELCAQWNRK 180
Db 121 ASGALNADGVADRTFTYIDPNNDIQFVSATAGSVGRNVEVLRVLDAQSDELCAQWNRK 180
Qy 181 GDDPTLDAGELLKASA 195
Db 181 GDDPTLDAGELLKASA 195

```

RESULT 9  
057529 MYCSM PRELIMINARY; PRT; 195 AA.

```

ID 057529_MYCSM PRELIMINARY; PRT; 195 AA.
AC 057529;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-FEB-2005 (TREMblrel. 29, Last annotation update)
DE Alkyl hydroperoxide reductase C.
GN Name=ahpc; Synonyms=Ahpc;
OS Mycobacterium smegmatis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxId=1772;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=mc2 155; PLASMID=pp81;
RX MEDLINE=96256622; PubMed=8655566;
RA Dhandayuthapani S., Zhang Y., Mudd M.H., Deretic V.;
RT "Oxidative stress response and its role in sensitivity to isoniazid in
RT Mycobacterium smegmatis and lack of expression in M. aurum and M.
RT tuberculosis."
RL J. Bacteriol. 178:3641-3649(1996).
DR EMBL; U43179; AAC44139.1; -; Genomic DNA.
DR EMBL; U31978; AAC44148.1; -; Genomic DNA.
DR HSSP; P32119; IQMV.
DR SMK; Q57529; 3-179.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR InterPro; IPR000866; Ahpc-TSA.
DR InterPro; IPR012336; Thioresdoxin-like.
DR InterPro; IPR012335; Thioresdoxin_fold.
DR Pfam; PF00578; Ahpc-TSA; 1.
KW Plasmid.
SQ SEQUENCE 195 AA; 21626 MW; 556A5C17A5DA937F CRC64;

```

Query Match 85.7%; Score 878; DB 2; Length 195;  
Best Local Similarity 82.2%; Pred. No. 5.8e-69;  
Matches 157; Conservative 20; Mismatches 14; Indels 0; Gaps 0;

```

Qy 1 MLLTTIGQPPAYQTLALIGDLSKVDAKOPGDYFTTTSDEHPGKRVVFWMPPKDFTFV 60
Db 1 MALLTIGQPPAYQTLALIGDLSKVDAKOPGDYFTTTSDEHPGKRVVFWMPPKDFTFV 60
Qy 61 CPTETAAFSKLNDEFERDAQILGVSIDSEFAHQWRAQHNDLKTLPFPMLSDIKRELSQ 120
Db 61 CPTETAAFGKLNDEFERDAQILGVSDNEFAHQWRAQHNDLKTLPFPMLSDIKRELSA 120
Qy 121 AAGVLNADGVADRTFTYIDPNNEIOFVSATAGSVGRNVDVLRVLDAQSDELCAQWNRK 180
Db 121 AAGVLNADGVADRTFTYIDPNNEIOFVSATAGSVGRNVDVLRVLDAQSDELCAQWNRK 180
Qy 181 GDDPTLDAGELLKASA 195
Db 181 GDDPTLDAGELLKASA 195

```

```

Qy 181 GDDPTLDAGELL 191
Db 181 GDDPTLDAGELL 191

```

RESULT 10  
06NGT3 CORDI PRELIMINARY; PRT; 198 AA.

```

ID 06NGT3_CORDI PRELIMINARY; PRT; 198 AA.
AC 06NGT3;
DT 05-JUL-2004 (TREMblrel. 27, Created)
DT 05-JUL-2004 (TREMblrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMblrel. 27, Last annotation update)
DE Iron repressible polypeptide (putative reductase).
GN Name=dtra; OrderedLocNames=DIP1420;
OS Corynebacterium diptheriae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxId=1717;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Biotype gravis / NCTC 13129;
RX MEDLINE=22965443; PubMed=14602910; DOI=10.1093/nar/gkg874;
RA Cerdano-Tarraga A.-M., Efstatiou A., Dover L.G., Holden M.T.G.,
RA Pallen M.J., Bentley S.D., Beara G.S., Churcher C.M., James K.D.,
RA De Zoya A., Chillingworth T., Cronin A., Dowd L., Fellwell T.,
RA Hamlin N., Holroyd S., Jagels K., Moule S., Quail M.A.,
RA Rabinowitsch E., Rutherford K.M., Thomson N.R., Unwin L.,
RA Whitehead S., Barrett B.G., Parkhill J.;
RT "The complete genome sequence and analysis of Corynebacterium
RT diptheriae NCTC13129."
RL Nucleic Acids Res. 31:6516-6523(2003).
DR EMBL; BX248358; CA849951.1; -; Genomic DNA.
DR InterPro; IPR000866; Ahpc-TSA.
DR InterPro; IPR012336; Thioresdoxin-like.
DR Pfam; PF00578; Ahpc-TSA; 1.
KW Complete proteome.
SQ SEQUENCE 198 AA; 22365 MW; 7A09C8F1B43205DD CRC64;

```

Query Match 67.7%; Score 693; DB 2; Length 198;  
Best Local Similarity 67.9%; Pred. No. 1.1e-52;  
Matches 125; Conservative 29; Mismatches 30; Indels 0; Gaps 0;

```

Qy 1 MLLTTIGQPPAYQTLALIGDLSKVDAKOPGDYFTTTSDEHPGKRVVFWMPPKDFTFV 60
Db 1 MSLLTIGQPPAYQTLALIGDLSKVDAKOPGDYFTTTSDEHPGKRVVFWMPPKDFTFV 60
Qy 61 CPTETAAFSKLNDEFERDAQILGVSIDSEFAHQWRAQHNDLKTLPFPMLSDIKRELSQ 120
Db 61 CPTETAAFGKLNDEFERDAQILGVSDNEFAHQWRAQHNDLKTLPFPMLSDIKRELSA 120
Qy 121 AAGVLNADGVADRTFTYIDPNNEIOFVSATAGSVGRNVDVLRVLDAQSDELCAQWNRK 180
Db 121 AAGVLNADGVADRTFTYIDPNNEIOFVSATAGSVGRNVDVLRVLDAQSDELCAQWNRK 180
Qy 181 GDDPTLDAGELL 191
Db 181 GDDPTLDAGELL 191

```

RESULT 11  
Q46025 CORDI PRELIMINARY; PRT; 198 AA.

```

ID 046025_CORDI PRELIMINARY; PRT; 198 AA.
AC 046025;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Iron repressible polypeptide.
GN Name=dtra;
OS Corynebacterium diptheriae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxId=1717;
RN [1]

```

RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C7;  
 RX MEDLINE=95286547; PubMed=7768861;  
 RA Tai S.S., Zhu Y.Y.;  
 RT "Cloning of a Corynebacterium diphtheriae iron-repressible gene that  
 shares sequence homology with the Ahpc subunit of alkyl hydroperoxide  
 reductase of *Salmonella typhimurium*.";  
 RL J. Bacteriol. 177:3512-3517(1995).  
 DR EMBL; U18620; AAA96946.1; -; Genomic\_DNA.  
 DR HSSP; Q9TX22; IE2Y.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR InterPro; IPR000866; Ahpc-TSA.  
 DR InterPro; IPR012336; Thioresdoxin-like.  
 DR InterPro; IPR012335; Thioresdoxin\_fold.  
 DR Pfam; PF00578; Ahpc-TSA; 1.  
 SQ SEQUENCE 198 AA; 22312 MW; 7A09DCBA07705DD CRC64;

Query Match 66.9%; Score 685; DB 2; Length 198;  
 Best Local Similarity 67.4%; Pred. No. 5.3e-52;  
 Matches 124; Conservative 29; Mismatches 31; Indels 0; Gaps 0;

QY 1 MULTIGDQPAAYQTLALIGDLSKVDAKOPGDYFTTTSDEHCKMVRVFFMPDPFTVC 60  
 DB 1 MSILTVGKPEPEFNLTKGSDLDHVNASQPEDYFETSLDKYEGKMKVVFYPDPFTVC 60  
 QY 61 CPTETAAFSKLNDEFEDRDAQILGVSIDSEFAHFQWRAQNDLKTLPFPMLSDIKRELISQA 120  
 DB 61 CPTETAAFGKLNDEFEDRDAQILGVSIDSEFAHFQWRAQNDLKTLPFPMLSDIKRELISQA 120  
 QY 121 AAGTLMADGVADRVTFIVDPNNETQFVSATAGSVGRNDEVLRVLDALQSEDLCAKMRK 180  
 DB 121 ALGVENEGVADRVATFIIDPGIIQFVSVTPDAGVNCNDEVLRVLDALQSEDLCAKMRK 180  
 QY 181 GDPPT 184  
 DB 181 NDPPT 184

RESULT 12  
 ID Q9XSV0\_STRVD PRELIMINARY; PRT; 194 AA.  
 AC Q9XSV0;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Alkyl hydroperoxide reductase.  
 OS Name=ahpc;  
 GN Streptomyces viridosporus.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=67581;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=77A;  
 RX MEDLINE=20356715; PubMed=10902909;  
 RA Ramachandran S., Magnuson T.S., Crawford D.L.;  
 RT "Isolation and analysis of three peroxide sensor regulatory gene  
 homologs ahpc, ahpx and oxvr in Streptomyces viridosporus 77A - a  
 lignocellulose degrading actinomycete";  
 RL DNA Seq. 11:51-60(2000).  
 DR EMBL; AF127576; AAD33340.1; -; Genomic\_DNA.  
 DR HSSP; Q9TX22; IE2Y.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0004601; F:peroxidase activity; IEA.  
 DR InterPro; IPR000866; Ahpc-TSA.  
 DR InterPro; IPR012336; Thioresdoxin-like.  
 DR InterPro; IPR012335; Thioresdoxin\_fold.  
 DR Pfam; PF00578; Ahpc-TSA; 1.  
 SQ SEQUENCE 194 AA; 21700 MW; F324A03842BC7AD1 CRC64;

Query Match 60.4%; Score 619; DB 2; Length 194;  
 Best Local Similarity 62.1%; Pred. No. 3.2e-46;

Matches 118; Conservative 18; Mismatches 46; Indels 8; Gaps 1;

QY 2 PLTTIGDQPAAYQTLALIGDLSKVDAKOPGDYFTTTSDEHCKMVRVFFMPDPFTVC 61  
 DB 10 PVLTVGDKPEPEFNLTKGSDLDHVNASQPEDYFETSLDKYEGKMKVVFYPDPFTVC 61  
 QY 62 PTEIAAFSKLNDEFEDRDAQILGVSIDSEFAHFQWRAQNDLKTLPFPMLSDIKRELISQA 121  
 DB 62 PTEIAAFGKLNDEFEDRDAQILGVSIDSEFAHFQWRAQNDLKTLPFPMLSDIKRELISQA 121  
 QY 122 AGVLMADGVADRVTFIVDPNNETQFVSATAGSVGRNDEVLRVLDALQSEDLCAKMRK 181  
 DB 122 LGIEGDEGFAKRAVTFIVDPNNETQFVSATAGSVGRNDEVLRVLDALQSEDLCAKMRK 181  
 QY 182 DPTLDAGELL 191  
 DB 182 DETIDPVALL 191

RESULT 13  
 ID Q9FBP5\_STRCO PRELIMINARY; PRT; 184 AA.  
 AC Q9FBP5;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Alkyl hydroperoxide reductase.  
 GN OrderedlocusNames=SC05032; ORFNames=SKC7.05c;  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteridae; Actinobacteriales; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.B., Quail M.A., Kleser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornby T., Howarth S.,  
 RA Huang C.-H., Kleser T., Larke L., Murphy L.D., Oliver K., O'Neill S.,  
 RA Rabinowitch E., Rajandream M.A., Rutherford K.M., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J.R., Bartell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2)";  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL339122; GAC05877.1; -; Genomic\_DNA.  
 DR HSSP; P32119; IQMV.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0004601; F:peroxidase activity; IEA.  
 DR InterPro; IPR000866; Ahpc-TSA.  
 DR InterPro; IPR012336; Thioresdoxin-like.  
 DR InterPro; IPR012335; Thioresdoxin\_fold.  
 DR Pfam; PF00578; Ahpc-TSA; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 184 AA; 20726 MW; F77E2820DA85E341 CRC64;

Query Match 57.9%; Score 593; DB 2; Length 184;  
 Best Local Similarity 61.4%; Pred. No. 5.3e-44;  
 Matches 116; Conservative 14; Mismatches 51; Indels 8; Gaps 1;

QY 3 LPLTIGDQPAAYQTLALIGDLSKVDAKOPGDYFTTTSDEHCKMVRVFFMPDPFTVC 62  
 DB 1 MLTVGDKPEPEFNLTKGSDLDHVNASQPEDYFETSLDKYEGKMKVVFYPDPFTVC 62  
 QY 63 TEIAAFSKLNDEFEDRDAQILGVSIDSEFAHFQWRAQNDLKTLPFPMLSDIKRELISQA 122  
 DB 53 TEIAAFGKLNDEFEDRDAQILGVSIDSEFAHFQWRAQNDLKTLPFPMLSDIKRELISQA 122  
 QY 123 GVLNADGVADRVTFIVDPNNETQFVSATAGSVGRNDEVLRVLDALQSEDLCAKMRK 182  
 DB 123 GVLNADGVADRVTFIVDPNNETQFVSATAGSVGRNDEVLRVLDALQSEDLCAKMRK 182

```

Db      113 GIEGEDGFAQRAVFIVDONNEIQFTMTAGSVGNPKSVLRVLDALQTDDELCPCKMSKGD 172
Qy      183 PTLDAGELL 191
Db      173 ETLDPVALL 181

RESULT 14
09RN72_STRCO
ID 09RN72_STRCO PRELIMINARY; PRT; 184 AA.
AC 09RN72_
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Ahpc.
GN Name=ahpc;
OS Streptomyces coelicolor A3(2).
OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
CX NCBI_TaxID=100226;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=A3;
RX MEDLINE=22206464; PubMed=12218006;
RX DOI=10.1128/JB.184.19.5214-5222.2002;
RA Hahn J.S., Oh S.Y., Roe J.H.;
RT "Role of OxyR as a peroxide-sensing positive regulator in Streptomyces
RT coelicolor A3(2).";
RN [2]
RN J. Bacteriol. 184:5214-5222(2002).
RL NUCLEOTIDE SEQUENCE.
RP STRAIN=A3;
RC Hahn J.-S., Roe J.-H.;
RA Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
RL EMBL; AF186371; AAF06744.1; -; Genomic_DNA.
DR HSSP; P32119; IQNV.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0004601; P:peroxidase activity; IEA.
DR InterPro; IPR000866; Ahpc-TSA.
DR InterPro; IPR012335; Thioresdoxin-like.
DR InterPro; IPR012335; Thioresdoxin_fold.
DR Pfam; PF00578; Ahpc-TSA; 1.
SQ SEQUENCE 184 AA; 20712 MW; E291762EC035E2EA CRC64;

Query Match 57.8%; Score 592; DB 2; Length 184;
Best Local Similarity 61.4%; Pred. No. 7.2e-44;
Matches 116; Conservative 14; Mismatches 51; Indels 8; Gaps 1;

Qy      3 LTTIGDQFPAYQTLALIGDLSKYDAKQPGDYFTTITSDEHPGKRVVFFMPKQFTFVCP 62
Db      1 MLTVGDKRPFEPDLTACV-----SLEKGDFOQINHKTVEGQKVFAMPKQFTFVCP 52

Qy      63 TEIAAFSKLNDPEFDRDAQIIIGVSIIDSEFAHFQWRQAQNDLKTLPFPMLSDIKELSGAA 122
Db      53 TEIAAFGLNDEPDRDAQIIIGVSGDSEFVHNAWKCHDRLDLPFPMIADSKHELMRDL 112

Qy      123 GVINADGVADRVTFIVDPNNEIQFVSATAGSVGNRVNDEVLRVLDALQSDDELCAQNRKGD 182
Db      113 GIEGEDGFAQRAVFIVDONNEIQFTMTAGSVGNPKSVLRVLDALQTDDELCPCKMSKGD 172

Qy      183 PTLDAGELL 191
Db      173 ETLDPVALL 181

RESULT 15
0821C6_STRAW
ID 0821C6_STRAW PRELIMINARY; PRT; 184 AA.
AC 0821C6_
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Putative alkyl hydroperoxide reductase.

```

```

GN Name=ahpc; OrderedLocusNames=SAV3232;
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
CX NCBI_TaxID=33903;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562; DOI=10.1038/nbt820;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T., Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C., Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T., Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR HSSP; BA000030; BAC70943.1; -; Genomic_DNA.
DR HSSP; Q9TXZ2; IEZY.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0004601; P:peroxidase activity; IEA.
DR InterPro; IPR000866; Ahpc-TSA.
DR InterPro; IPR012336; Thioresdoxin-like.
DR InterPro; IPR012335; Thioresdoxin_fold.
DR Pfam; PF00578; Ahpc-TSA; 1.
KW Complete proteome.
SQ SEQUENCE 184 AA; 20773 MW; 6F98071259D81088 CRC64;

Query Match 57.6%; Score 590; DB 2; Length 184;
Best Local Similarity 59.3%; Pred. No. 1.1e-43;
Matches 112; Conservative 21; Mismatches 48; Indels 8; Gaps 1;

Qy      3 LTTIGDQFPAYQTLALIGDLSKYDAKQPGDYFTTITSDEHPGKRVVFFMPKQFTFVCP 62
Db      1 MLTVGDKRPFEPDLTACV-----SLEKGNFQQINHKTVEGQKVFAMPKQFTFVCP 52

Qy      63 TEIAAFSKLNDPEFDRDAQIIIGVSIIDSEFAHFQWRQAQNDLKTLPFPMLSDIKELSGAA 122
Db      53 TEIAAFGLNDEPDRDAQIIIGVSGDSEFVHNAWKCHDRLDLPFPMIADSKHELMRDL 112

Qy      123 GVINADGVADRVTFIVDPNNEIQFVSATAGSVGNRVNDEVLRVLDALQSDDELCAQNRKGD 182
Db      113 GIEGEDGFAQRAVFIVDONNEIQFVSATAGSVGNPKSVLRVLDALQTDDELCPCKMSKGD 172

Qy      183 PTLDAGELL 191
Db      173 ETLDPVALL 181

Search completed: March 23, 2006, 05:24:25
Job time : 50.8737 secs

```

This Page Blank (uspto)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using SW model

Run on: March 23, 2006, 05:24:53 ; Search time 13.2997 Seconds  
(without alignments)  
1212.194 Million cell updates/sec

Title: US-10-617-038-26

Perfect score: 1024

Sequence: 1 MPLLTIGDPAYQLTALIG.....GNWRKDPYLDAGELLKASA 195

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/1aa/5\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/6\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/H\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/PCRTS\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/RE\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	898	87.7	195	2	US-08-311-731A-84 Sequence 84, Appl
2	524.5	51.2	184	2	US-09-902-540-10665 Sequence 10665, A
3	350.5	34.2	982	2	US-09-551-974A-95 Sequence 95, Appl
4	350.5	34.2	982	2	US-09-565-501A-95 Sequence 95, Appl
5	350.5	34.2	982	2	US-09-639-206A-95 Sequence 95, Appl
6	350.5	34.2	982	2	US-09-874-923-95 Sequence 95, Appl
7	350.5	34.2	1427	2	US-09-551-974A-97 Sequence 97, Appl
8	350.5	34.2	1427	2	US-09-565-501A-97 Sequence 97, Appl
9	350.5	34.2	1427	2	US-09-639-206A-97 Sequence 97, Appl
10	350.5	34.2	1427	2	US-09-874-923-97 Sequence 97, Appl
11	350.5	34.2	1464	2	US-10-012-886-1008 Sequence 1008, Ap
12	350.5	34.2	1641	2	US-09-551-974A-96 Sequence 96, Appl
13	350.5	34.2	1641	2	US-09-565-501A-96 Sequence 96, Appl
14	350.5	34.2	1641	2	US-09-639-206A-96 Sequence 96, Appl
15	350.5	34.2	1641	2	US-09-874-923-96 Sequence 96, Appl
16	346.5	33.8	199	2	US-09-183-861-24 Sequence 24, Appl
17	346.5	33.8	199	2	US-09-022-765-24 Sequence 24, Appl
18	346.5	33.8	199	2	US-09-551-974A-24 Sequence 24, Appl
19	346.5	33.8	199	2	US-09-565-501A-24 Sequence 24, Appl
20	346.5	33.8	199	2	US-09-639-206A-24 Sequence 24, Appl
21	346.5	33.8	199	2	US-09-874-923-24 Sequence 24, Appl
22	346.5	33.8	199	2	US-08-798-841-24 Sequence 24, Appl
23	344	33.6	206	2	US-09-183-861-26 Sequence 26, Appl
24	344	33.6	206	2	US-09-022-765-26 Sequence 26, Appl
25	344	33.6	206	2	US-09-551-974A-26 Sequence 26, Appl
26	344	33.6	206	2	US-09-565-501A-26 Sequence 26, Appl
27	344	33.6	206	2	US-09-639-206A-26 Sequence 26, Appl

28	344	33.6	206	2	US-09-874-923-26 Sequence 26, Appl
29	344	33.6	206	2	US-08-798-841-26 Sequence 26, Appl
30	327.5	32.0	244	2	US-09-252-991A-29692 Sequence 29692, A
31	323.5	31.6	205	2	US-09-489-039A-13084 Sequence 13084, A
32	320.5	31.3	199	1	US-08-467-265-17 Sequence 17, Appl
33	320.5	31.3	199	1	US-08-467-265-17 Sequence 17, Appl
34	320.5	31.3	199	2	US-09-407-891-17 Sequence 1055, Ap
35	318	31.1	206	2	US-09-538-092-1055 Sequence 8377, Ap
36	318	31.1	257	1	US-09-949-016-8377 Sequence 16, Appl
37	318	31.1	257	1	US-08-467-265-16 Sequence 16, Appl
38	318	31.1	257	2	US-08-467-265-16 Sequence 16, Appl
39	318	31.1	257	2	US-09-407-891-16 Sequence 16, Appl
40	318	31.1	257	2	US-09-375-907-5 Sequence 5, Appl1
41	317	31.0	248	2	US-09-270-767-46784 Sequence 46784, A
42	317	31.0	256	2	US-09-538-092-1039 Sequence 1039, Ap
43	317	31.0	256	2	US-09-949-016-6760 Sequence 6760, Ap
44	317	31.0	258	2	US-09-949-016-10249 Sequence 10249, A
45	315.5	30.8	199	2	US-09-375-907-4 Sequence 4, Appl1

## ALIGNMENTS

RESULT 1  
US-08-311-731A-84  
; Sequence 84, Application US/08311731A  
; Patent No. 6583266  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, DOUGLAS  
; APPLICANT: MAO, JEN-I  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 411  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P. C.  
; STREET: 600 ATLANTIC AVENUE  
; CITY: BOSTON  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 02210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/311, 731A  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GATES, EDWARD R.  
; REGISTRATION NUMBER: 31,616  
; REFERENCE/DOCKET NUMBER: C0044/7125  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/720-3500  
; TELEFAX: 617/720-2441  
; INFORMATION FOR SEQ ID NO: 84:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 195 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: MYCOBACTERIUM LAPRAE  
; US-08-311-731A-84  
Query Match 87.7%; Score 898; DB 2; Length 195;  
Best Local Similarity 85.6%; Pred. No. 1.6e-100;  
Matches 167; Conservative 14; Mismatches 0; Gaps 0;  
QY 1 MPLLTIGDPAYQLTALIGDLSKVDKAKQGDYFTTTSDEHFGKMRVVFPMKDPFTFV 60



```
Db 1 MSLSIGQFPAYQTLALIGDLSKVDAAQPDYFTTTSDEHPGKRVVFPMPKDPFTFPI 60
Qy 61 CPTTEAASKLNDEFEEDBDQAQILGVSIDSEFAHFQWRAQHNDLKTLPPEMLSDIKRELSQ 120
Db 61 CPTTEAAGKLNDEFEEDBDQAQILGVSIDSEFAHFQWRAQHNDLKTLPPEMLSDIKRELSQ 120
Qy 121 AAGVADGVADRVTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDLCACNNRK 180
Db 121 AAGVADGVADRVTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDLCACNNRK 180
Qy 181 GDPTLADGELLKASA 195
Db 181 GDPTLADGELLKASA 195
```

RESULT 2  
US-09-902-540-10665  
Sequence 10665, Application US/09902540  
Patent No. 6833447

GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15849)B  
CURRENT FILING DATE: US/09/902,540  
CURRENT APPLICATION NUMBER: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 10665  
LENGTH: 184  
TYPE: PRT  
ORGANISM: Myxococcus xanthus  
US-09-902-540-10665

Query Match 51.2%; Score 524.5; DB 2; Length 184;  
Best Local Similarity 54.3%; Pred. No. 3e-55;  
Matches 100; Conservative 25; Mismatches 50; Indels 9; Gaps 2;

```
Qy 3 LITIGDQFPAYQTLALIGDLSKVDAAQPDYFTTTSDEHPGKRVVFPMPKDPFTFVCP 62
Db 1 MLTIVGDKIPNKFKVATV-----SLEKGEFODITETFGKMLVLFAMPKDPFTFICP 52
Qy 63 TEIAFSKLNDEFEEDBDQAQILGVSIDSEFAHFQWRAQHNDLKTLPPEMLSDIKRELSQAA 122
Db 53 TEIAFSGKLNDEFEEDBDQAQILGVSIDSEFAHFQWRAQHNDLKTLPPEMLSDIKRELSQAA 112
Qy 123 GVLA-ADGVADRVTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDLCACNNRK 181
Db 113 GILKHEBVALRAITFIADPEGIIHVTYNDLSVGRNVSETIRTDALQTDLCACNNRK 172
Qy 182 DPTL 185
Db 173 ERTL 176
```

RESULT 3  
US-09-551-974A-95  
Sequence 95, Application US/09551974A  
Patent No. 6500437  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, David C.  
APPLICANT: Skeiky, Yasir A.W.  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
FILE REFERENCE: 210121.420C5  
CURRENT APPLICATION NUMBER: US/09/551.974A

CURRENT FILING DATE: 2000-04-14  
NUMBER OF SEQ ID NOS: 101  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 95  
LENGTH: 982  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Fusion (poly-protein) constructs comprising multiple  
US-09-551-974A-95

Query Match 34.2%; Score 350.5; DB 2; Length 982;  
Best Local Similarity 39.4%; Pred. No. 5.2e-33;  
Matches 74; Conservative 35; Mismatches 60; Indels 19; Gaps 4;

```
Qy 6 IGQDQFPAYQTLALIGDLSKVDAAQPDYFTTTSDEHPGKRVVFPMPKDPFTFVCPTEI 65
Db 15 INSPAPSEFEVVALM-----PNSFKKISLSYKSKRWVLFYFIDFTFVCPTEV 63
Qy 66 AAFSKLNDEFEEDBDQAQILGVSIDSEFAHFQWRAQ---HNDLKTLPPEMLSDIKRELSQAA 122
Db 64 IAFSDSVSRFENLNCVELACISDEVAHLQMTLQDRKKGLGTMALPMLADTKTSIARSY 123
Qy 123 GVL-MADGVADRVTFIVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQ---SDELCACN 177
Db 124 GVLEESQGVAVYRGIFIDPHGMRLQITVNDMPVGRSVSEVLRILLEAFQVEKRGVCPAN 183
Qy 178 WRKGDPTL 185
Db 184 WKKGAPTM 191
```

RESULT 4  
US-09-565-501A-95  
Sequence 95, Application US/09565501A  
Patent No. 6607731

GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, David C.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Bhatia, Ajay  
APPLICANT: Coler, Rhea  
APPLICANT: Peter Probst  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
FILE REFERENCE: 210121.420C6  
CURRENT APPLICATION NUMBER: US/09/565.501A  
CURRENT FILING DATE: 2000-05-05  
NUMBER OF SEQ ID NOS: 112  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 95  
LENGTH: 982  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Fusion (poly-protein) constructs comprising multiple  
US-09-565-501A-95

Query Match 34.2%; Score 350.5; DB 2; Length 982;  
Best Local Similarity 39.4%; Pred. No. 5.2e-33;  
Matches 74; Conservative 35; Mismatches 60; Indels 19; Gaps 4;

```
Qy 6 IGQDQFPAYQTLALIGDLSKVDAAQPDYFTTTSDEHPGKRVVFPMPKDPFTFVCPTEI 65
Db 15 INSPAPSEFEVVALM-----PNSFKKISLSYKSKRWVLFYFIDFTFVCPTEV 63
Qy 66 AAFSKLNDEFEEDBDQAQILGVSIDSEFAHFQWRAQ---HNDLKTLPPEMLSDIKRELSQAA 122
Db 64 IAFSDSVSRFENLNCVELACISDEVAHLQMTLQDRKKGLGTMALPMLADTKTSIARSY 123
```

[illegible]

RESULT 5  
US-09-639-206A-95  
; Sequence 95, Application US/09639206A  
; Patent No. 6612227

```

/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Campos-Neto, Antonio
/ APPLICANT: Webb, John R.
/ APPLICANT: Dillion, David C.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Bhalla, Ajay
/ APPLICANT: Cole, Rhea
/ APPLICANT: Probst, Peter
/ TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
/ TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
/ FILE REFERENCE: 210121.420C7
/ CURRENT APPLICATION NUMBER: US/09/639,206A
/ CURRENT FILING DATE: 2000-08-14
/ NUMBER OF SEQ. ID NOS: 112
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 95
/ LENGTH: 982
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Fusion (poly-protein) constructs comprising multiple
/ OTHER INFORMATION: Leishmania antigens
/ US-09-639-206A-95

```

	Query Match	39.4%;	Score 350.5;	DB 2;	length 982;	
	Best Local Similarity	39.4%;	Pred. No. 5.2e-33;			
	Matches	74;	Conservative	35;	Mismatches	60; Indels 19; Gaps 4
QY	6	IGDQPAYOLITALIGSDLSKVADKQBGDYFTTITSDEHPGKRVRVFVMPKDFTVCPTETI	65			
	:::	:::	:::	:::	:::	:::
Db	15	INSPAPSFEEVALM-----PNCGFKKISLSSYGKKVVLFPPPLDFTFVCETE	63			
QY	66	AASKSUNDEBEDDAOILGVSIDSEPAHFCOMRAO---HNDDKTLPFPMLSDIKRELSOA	122			
		:::			:	
Db	64	IATSDSVSRNEINCEVLACSIDSEPAHLQMTLQRRKGGLGTMAIPMLADKRTSINSY	123			
QY	123	GVL-NMDGVADRVTVPDPNRIQFVSATAGSVGRNVDEVLRVDALQ----SDELCA	177			
	:::	:::	:::	:::	:::	:::
Db	124	GVLEESGVAVRGLFIIDPHGMRLQTIVNDMVGRSVSEVLRLLAEAFGEVKHGVSVC	183			
QY	178	WRKGDPTL 185				
	:::	:::	:::	:::	:::	:::
Db	184	WKKGAPTM 191				

RESULT 6  
US-09-874-923-95  
Sequence 95, Application US/09874922.1  
Patent No. 6638517  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, David C.  
APPLICANT: Skelley, Yair A.W.  
APPLICANT: Bhatia, Ajay  
APPLICANT: Colter, Rhea  
APPLICANT: Probst, Peter

```

1  APPLICANT: Brannon, Mark
2  TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
3  TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
4  FILE REFERENCE: 210121.420C8
5  CURRENT APPLICATION NUMBER: US/09/874,923
6  CURRENT FILING DATE: 2001-06-04
7  NUMBER OF SEQ ID NOS: 122
8  SOFTWARE: FastSeq for Windows Version 4.0
9  SEQ ID NO 95
10 LENGTH: 982
11 TYPE: PRT
12 ORGANISM: Artificial Sequence
13 FEATURE:
14 OTHER INFORMATION: Fusion (poly-protein) constructs comprising multiple
15 OTHER INFORMATION: Leishmania antigens
16 US-09-874-923-95
17
18 Query Match      34.2%; Score 350.5; DB 2; Length 982;
19 Best Local Similarity 35.4%; Pred. No. 5.2e-33;
20 Matches 74; Conservative 35; Mismatches 60; Indels 19; Gaps 4;

```

Query Match	34.2%	Score 350.5	DB 2	Length 982
Best Local Similarity	39.4%	Pred. No. 5.2e-33		
Matches	74	Conservative 35	Mismatches 60	Indels 19
			Gaps	4
QY	6	IGDGFAYVQLTALIGDLSKYDAKQPGDYFTTISDSEHPGKMRVVFVFPKQFTVCPREI	65	
		:::		
Db	15	INSPAPSEFEEVALM-----PNSGSEFKSISLSYKGMVVLFFYPIDTFVCPTEV	63	
QY	66	AAFEKLNDEFEDRAOAILGVSIDSEFAHPQWRAQ---HNIDKLTPLFPPLTSDIKRELSDQA	122	
		:::		
Db	64	IAPFDSVSRRFELNCEVLACSIDSEYHNLQMTLDRKKGGIGLGTNAIPIPLADKTSIASY	123	
QY	123	GVL-NADGVADRYFTIVDPNNNEIQFSATAGSVGRNVDYLIVDALQ---SDPLCAQN	177	
		:::		
Db	124	GVLEESQGVAVRGFLITDPHGLMLRQITVNDMPVGRSVSEVRLILEAQFVEKHEGVECPAN	183	
QY	178	WRKGDPTL	185	
		:::		
Db	184	WRKGAPTM	191	

```

RESULT 7
US-09-551-974A-97
/ Sequence 97, Application US/09551974A
/ Patent No. 6500437
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Campos-Neto, Antonio
/ APPLICANT: Webb, John R.
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Skelly, Yasir A.W.
/ TITLE OF INVENTION: LELISHMANIA ANTIGENS FOR USE IN THE
/ TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LELISHMANIASIS
/ FILE REFERENCE: 210121.420C5
/ CURRENT APPLICATION NUMBER: US/09/551,974A
/ CURRENT FILING DATE: 2000-04-14
/ NUMBER OF SEQ ID NOS: 101
/ SOFTWARE: FaSTSeq for Windows Version 4.0
/ SEQ ID NO 97
/ LENGTH: 1427
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Fusion (poly-protein) constructs comprising multiple
/ OTHER INFORMATION: Leishmania antigens
US-09-551-974A-97

```

Query Match 34.2%; Score 350.5; DB 2; Length 1427;  
 Best Local Similarity 39.4%; Pred. No. 9.3e-33;  
 Matches 74; Conservative 35; Mismatches 60; Indels 19; Gaps 4;

```

QY      6  IGDPPAYQVLITGSLSKYDAKQPDYFTTISDSDPGKRVVFEMPKDFTVCPREI 65
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      15  INSPAPFEFEVALM-----PQSGFKLISLSSYKGRVVVLFYPLDFTVCPREV 63
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      66  AAFSKLNDFFBDRDAQLGVGSIDSEFAHFMORAO--HNDLKLTPFPMLSIDIKELSLQA 122
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      66  AAFSKLNDFFBDRDAQLGVGSIDSEFAHFMORAO--HNDLKLTPFPMLSIDIKELSLQA 122
  
```

```
Db      64 IAFSSVSFRFNEELNCEVLACSIDSEYAHLOQTLDQRKKGGIGCTMAIPMLADTKSIASYS 123
      123 GVL-NADGVADRVTFIYDPNNEIOFVSATAGSVGRNDEVLRVLDALQ----SDELCACN 177
      124 GVLEESQGVAVRGLFIIDPHGMLRQITVNDMPVGRSVSEVLRLLAEAFQVFKGGEVCPAN 183
Qy      178 WRKGDPTL 185
      184 WKKGAPTM 191
```

RESULT 8  
US-09-565-501A-97  
Sequence 97, Application US/09565501A  
Patent No. 6607731

GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Bhatia, Ajay  
APPLICANT: Coler, Rhea  
APPLICANT: Peter Probst  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS  
FILE REFERENCE: 210121.420C6  
CURRENT APPLICATION NUMBER: US/09/565,501A  
CURRENT FILING DATE: 2000-05-05  
NUMBER OF SEQ ID NOS: 112  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 97  
LENGTH: 1427  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Fusion (poly-protein) constructs comprising multiple  
US-09-565-501A-97

Query Match 34.2%; Score 350.5; DB 2; Length 1427;  
Best Local Similarity 39.4%; Pred. No. 9.3e-33;  
Matches 74; Conservative 35; Mismatches 60; Indels 19; Gaps 4;  
Qy 6 IGDQPAYQLTALIGDSLKVDKQPGDYFTTITSDSHPGKRVVFPMKDFVPCPTETI 65  
Db 15 INSPAPSEEVYALM-----PNSFKKISLSYKGMVLFYFPDFTVCPTEV 63  
Qy 66 AAFSKLDEFEDRDQAIIIGVSIDSEFAHFQWRAQ---HNDLKTLPFPMLSDIKRELQAA 122  
Db 64 IAFDSVSFRFNEELNCEVLACSIDSEYAHLOQTLDQRKKGGIGCTMAIPMLADTKSIASYS 123  
Qy 123 GVL-NADGVADRVTFIYDPNNEIOFVSATAGSVGRNDEVLRVLDALQ----SDELCACN 177  
Db 124 GVLEESQGVAVRGLFIIDPHGMLRQITVNDMPVGRSVSEVLRLLAEAFQVFKGGEVCPAN 183  
Qy 178 WRKGDPTL 185  
Db 184 WKKGAPTM 191

RESULT 9  
US-09-639-206A-97  
Sequence 97, Application US/09639206A  
Patent No. 6613337  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Bhatia, Ajay

APPLICANT: Coler, Rhea  
APPLICANT: Probst, Peter  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS  
FILE REFERENCE: 210121.420C7  
CURRENT APPLICATION NUMBER: US/09/639,206A  
CURRENT FILING DATE: 2000-08-14  
NUMBER OF SEQ ID NOS: 112  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 97  
LENGTH: 1427  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Fusion (poly-protein) constructs comprising multiple  
US-09-639-206A-97

Query Match 34.2%; Score 350.5; DB 2; Length 1427;  
Best Local Similarity 39.4%; Pred. No. 9.3e-33;  
Matches 74; Conservative 35; Mismatches 60; Indels 19; Gaps 4;  
Qy 6 IGDQPAYQLTALIGDSLKVDKQPGDYFTTITSDSHPGKRVVFPMKDFVPCPTETI 65  
Db 15 INSPAPSEEVYALM-----PNSFKKISLSYKGMVLFYFPDFTVCPTEV 63  
Qy 66 AAFSKLDEFEDRDQAIIIGVSIDSEFAHFQWRAQ---HNDLKTLPFPMLSDIKRELQAA 122  
Db 64 IAFDSVSFRFNEELNCEVLACSIDSEYAHLOQTLDQRKKGGIGCTMAIPMLADTKSIASYS 123  
Qy 123 GVL-NADGVADRVTFIYDPNNEIOFVSATAGSVGRNDEVLRVLDALQ----SDELCACN 177  
Db 124 GVLEESQGVAVRGLFIIDPHGMLRQITVNDMPVGRSVSEVLRLLAEAFQVFKGGEVCPAN 183  
Qy 178 WRKGDPTL 185  
Db 184 WKKGAPTM 191

RESULT 10  
US-09-874-923-97  
Sequence 97, Application US/09874923  
Patent No. 6638517  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Bhatia, Ajay  
APPLICANT: Coler, Rhea  
APPLICANT: Peter Probst  
APPLICANT: Brannon, Mark  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS  
FILE REFERENCE: 210121.420C8  
CURRENT APPLICATION NUMBER: US/09/874,923  
CURRENT FILING DATE: 2001-06-04  
NUMBER OF SEQ ID NOS: 122  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 97  
LENGTH: 1427  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Fusion (poly-protein) constructs comprising multiple  
US-09-874-923-97

Query Match 34.2%; Score 350.5; DB 2; Length 1427;  
Best Local Similarity 39.4%; Pred. No. 9.3e-33;  
Matches 74; Conservative 35; Mismatches 60; Indels 19; Gaps 4;



```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 96
; LENGTH: 1641
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion (poly-protein) constructs comprising multiple
; OTHER INFORMATION: Leishmania antigens
US-09-565-501A-96
```

```

Query Match          34.2%; Score 350.5; DB 2; Length 1641;
Best Local Similarity 39.4%; Pred. No. 1.2e-32;
Matches 74; Conservative 35; Mismatches 60; Indels 19; Gaps 4;
```

```

QY 6 IGDQFPAYQLTALIGDLSKVDKQPGDYFTTITSDEHFGKRVVFFWPKDFTFVCPTEI 65
DB 15 INSPAPSEFEVYALM-----PNSGFKKISLSSTYKGMVVLFFYPPLDFTFVCPTEV 63
QY 66 AAFSKLNDDEFDRDAQILGVSIIDSEFAHFQWRAQ---HNDLKTLPFPMLSDIKRELSQA 122
DB 64 IAFSDSVSRFVELNCEVLACSIDSEYAHLOWTLDQRKKGGIGTWAIPMLADTKTSIARSY 123
QY 123 GVL-NADGVADRVTFIYDPNNEIQFVSATAGSVGRNDEVLRVLDALQ---SDELCAQN 177
DB 124 GVLSESGVAVRGFLFIIDPHGMLRQITVNDMPVGRSVBEVLRLLEAFQFVEKHGEVCPAN 183
QY 178 WRKGDPPTL 185
DB 184 WKKGAPTM 191
```

```

RESULT 14
US-09-639-206A-96
; Sequence 96, Application US/09639206A
; Patent No. 6613337
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neco, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, David C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Bhactia, Ajay
; APPLICANT: Colier, Rhea
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; FILE REFERENCE: 210121.420C7
; CURRENT APPLICATION NUMBER: US/09/639,206A
; CURRENT FILING DATE: 2000-08-14
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 96
; LENGTH: 1641
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion (poly-protein) constructs comprising multiple
; OTHER INFORMATION: Leishmania antigens
US-09-639-206A-96
```

```

Query Match          34.2%; Score 350.5; DB 2; Length 1641;
Best Local Similarity 39.4%; Pred. No. 1.2e-32;
Matches 74; Conservative 35; Mismatches 60; Indels 19; Gaps 4;
```

```

QY 6 IGDQFPAYQLTALIGDLSKVDKQPGDYFTTITSDEHFGKRVVFFWPKDFTFVCPTEI 65
DB 15 INSPAPSEFEVYALM-----PNSGFKKISLSSTYKGMVVLFFYPPLDFTFVCPTEV 63
QY 66 AAFSKLNDDEFDRDAQILGVSIIDSEFAHFQWRAQ---HNDLKTLPFPMLSDIKRELSQA 122
DB 64 IAFSDSVSRFVELNCEVLACSIDSEYAHLOWTLDQRKKGGIGTWAIPMLADTKTSIARSY 123
QY 123 GVL-NADGVADRVTFIYDPNNEIQFVSATAGSVGRNDEVLRVLDALQ---SDELCAQN 177
```

```

DB 124 GVLSESGVAVRGFLFIIDPHGMLRQITVNDMPVGRSVBEVLRLLEAFQFVEKHGEVCPAN 183
QY 178 WRKGDPPTL 185
DB 184 WKKGAPTM 191
```

```

RESULT 15
US-09-874-923-96
```

```

; Sequence 96, Application US/09874923
; Patent No. 6638517
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neco, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, David C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Bhactia, Ajay
; APPLICANT: Colier, Rhea
; APPLICANT: Probst, Peter
; APPLICANT: Brannon, Mark
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; FILE REFERENCE: 210121.420C8
; CURRENT APPLICATION NUMBER: US/09/874,923
; CURRENT FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 96
; LENGTH: 1641
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion (poly-protein) constructs comprising multiple
; OTHER INFORMATION: Leishmania antigens
US-09-874-923-96
```

```

Query Match          34.2%; Score 350.5; DB 2; Length 1641;
Best Local Similarity 39.4%; Pred. No. 1.2e-32;
Matches 74; Conservative 35; Mismatches 60; Indels 19; Gaps 4;
```

```

QY 6 IGDQFPAYQLTALIGDLSKVDKQPGDYFTTITSDEHFGKRVVFFWPKDFTFVCPTEI 65
DB 15 INSPAPSEFEVYALM-----PNSGFKKISLSSTYKGMVVLFFYPPLDFTFVCPTEV 63
QY 66 AAFSKLNDDEFDRDAQILGVSIIDSEFAHFQWRAQ---HNDLKTLPFPMLSDIKRELSQA 122
DB 64 IAFSDSVSRFVELNCEVLACSIDSEYAHLOWTLDQRKKGGIGTWAIPMLADTKTSIARSY 123
QY 123 GVL-NADGVADRVTFIYDPNNEIQFVSATAGSVGRNDEVLRVLDALQ---SDELCAQN 177
DB 124 GVLSESGVAVRGFLFIIDPHGMLRQITVNDMPVGRSVBEVLRLLEAFQFVEKHGEVCPAN 183
QY 178 WRKGDPPTL 185
DB 184 WKKGAPTM 191
```

```

Search completed: March 23, 2006, 05:30:15
Job time : 13.3997 secs
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using SW model

Run on: March 23, 2006, 06:14:33 ; Search time 41.0073 Seconds  
(without alignments)  
1986.885 Million cell updates/sec

Title: US-10-617-038-26

Perfect score: 1024  
Sequence: 1 MFLRTIGDQPAYQUTALIG.....CNMRKDPDLDAELLKASA 195

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA\_Main:\*  
1: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1024	100.0	195	4	US-10-282-122A-62742
2	1024	100.0	195	4	US-10-282-122A-64708
3	1024	100.0	195	4	US-10-617-038-26
4	1024	100.0	195	4	US-10-332-512A-21
5	1024	100.0	195	5	US-10-732-923-21281
6	943	97.1	195	5	US-10-732-923-21020
7	935	91.3	195	4	US-10-282-122A-61886
8	935	91.3	195	5	US-10-732-923-21016
9	920	89.8	195	5	US-10-732-923-21018
10	904	88.3	195	5	US-10-282-122A-63991
11	904	88.3	195	5	US-10-732-923-21017
12	878	85.7	195	5	US-10-732-923-21019
13	693	67.7	198	4	US-10-282-122A-53752
14	685	66.9	198	5	US-10-732-923-21013
15	619	60.4	194	5	US-10-732-923-21258
16	592	57.9	184	5	US-10-732-923-20950
17	592	57.8	184	5	US-10-732-923-20951
18	590	57.6	184	5	US-10-732-923-21095
19	589	57.5	184	4	US-10-156-761-10769
20	569.5	55.6	179	5	US-10-732-923-21094
21	530.5	51.8	188	4	US-10-369-493-19388
22	475	46.4	185	5	US-10-732-923-21306
23	453	43.2	184	5	US-10-732-923-21050
24	444	43.4	184	5	US-10-732-923-21087
25	444	43.4	184	5	US-10-732-923-21088
26	432.5	42.2	182	4	US-10-282-122A-49002
27	432.5	42.2	182	5	US-10-732-923-21322

28	429.5	41.9	230	5	US-10-732-923-21109	Sequence 21109, A
29	420.5	41.1	182	5	US-10-732-923-20992	Sequence 20992, A
30	410.5	40.1	182	4	US-10-282-122A-50913	Sequence 50913, A
31	390	36.1	223	5	US-10-732-923-21104	Sequence 21104, A
32	371	36.2	194	5	US-10-732-923-21229	Sequence 21229, A
33	361.5	35.3	180	4	US-10-369-493-17716	Sequence 17716, A
34	361.5	35.3	180	5	US-10-732-923-21084	Sequence 21084, A
35	360	35.2	199	5	US-10-732-923-21100	Sequence 21100, A
36	358.5	35.0	200	5	US-10-732-923-21268	Sequence 21268, A
37	353.5	34.5	193	4	US-10-369-493-20936	Sequence 20936, A
38	353.5	34.5	200	5	US-10-732-923-21284	Sequence 21284, A
39	352	34.4	197	5	US-10-732-923-21042	Sequence 21042, A
40	351.5	34.3	192	4	US-10-369-493-18827	Sequence 18827, A
41	351.5	34.3	199	5	US-10-732-923-21049	Sequence 21049, A
42	351.5	34.3	203	5	US-10-732-923-20977	Sequence 20977, A
43	351	34.3	199	5	US-10-732-923-21152	Sequence 21152, A
44	351	34.3	199	5	US-10-732-923-21153	Sequence 21153, A
45	351	34.3	199	5	US-10-732-923-21206	Sequence 21206, A

## ALIGNMENTS

RESULT 1  
US-10-282-122A-62742  
; Sequence 62742, Application US/10282122A  
; Publication No. US20040029129A1  
GENERAL INFORMATION:  
; APPLICANT: Wang, Liangu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zykied, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Foreyth, R.  
; APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282.122A  
PRIOR FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 62742  
LENGTH: 195  
TYPE: PRT  
ORGANISM: Mycobacterium bovis  
US-10-282-122A-62742

Query Match 100.0%; Score 1024; DB 4; Length 195;  
Best Local Similarity 100.0%; Pred. No. 3.2e-103;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPLLITGDOPFAVQULTALIGDLSKVDAKOPGDYFTTITSDHPGKRWVFFMPKDFTV 60  
Db 1 MPLLITGDOPFAVQULTALIGDLSKVDAKOPGDYFTTITSDHPGKRWVFFMPKDFTV 60

Qy 61 CPTETIAAFSKLNDEFEDRDAQILGVSIIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQ 120  
Db 61 CPTETIAAFSKLNDEFEDRDAQILGVSIIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQ 120

Qy 121 AAGVLANADGVADRVTFTVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQWNRK 180  
Db 121 AAGVLANADGVADRVTFTVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQWNRK 180

Qy 181 GDFPTLDAGELKASA 195  
Db 181 GDFPTLDAGELKASA 195

RESULT 2  
US-10-282-122A-64708  
; Sequence 64708, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See file wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 64708  
; LENGTH: 195  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-282-122A-64708

Query Match 100.0%; Score 1024; DB 4; Length 195;  
Best Local Similarity 100.0%; Pred. No. 3.2e-103;

Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPLLITGDOPFAVQULTALIGDLSKVDAKOPGDYFTTITSDHPGKRWVFFMPKDFTV 60  
Db 1 MPLLITGDOPFAVQULTALIGDLSKVDAKOPGDYFTTITSDHPGKRWVFFMPKDFTV 60

Qy 61 CPTETIAAFSKLNDEFEDRDAQILGVSIIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQ 120  
Db 61 CPTETIAAFSKLNDEFEDRDAQILGVSIIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQ 120

Qy 121 AAGVLANADGVADRVTFTVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQWNRK 180  
Db 121 AAGVLANADGVADRVTFTVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQWNRK 180

Qy 181 GDFPTLDAGELKASA 195  
Db 181 GDFPTLDAGELKASA 195

RESULT 3  
US-10-617-038-26  
; Sequence 26, Application US/10617038  
; Publication No. US20040057963A1  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Peter  
; APPLICANT: Stryhn, Anette  
; APPLICANT: Rosenkrands, Ida  
; TITLE OF INVENTION: Therapeutic TB Vaccine  
; FILE REFERENCE: S815AUSA  
; CURRENT APPLICATION NUMBER: US/10/617,038  
; CURRENT FILING DATE: 2003-07-11  
; PRIOR APPLICATION NUMBER: DK PA 2002 01098  
; PRIOR FILING DATE: 2002-07-13  
; PRIOR APPLICATION NUMBER: US 60/401,725  
; PRIOR FILING DATE: 2002-08-07  
; NUMBER OF SEQ ID NOS: 187  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 26  
; LENGTH: 195  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-617-038-26

Query Match 100.0%; Score 1024; DB 4; Length 195;  
Best Local Similarity 100.0%; Pred. No. 3.2e-103;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPLLITGDOPFAVQULTALIGDLSKVDAKOPGDYFTTITSDHPGKRWVFFMPKDFTV 60  
Db 1 MPLLITGDOPFAVQULTALIGDLSKVDAKOPGDYFTTITSDHPGKRWVFFMPKDFTV 60

Qy 61 CPTETIAAFSKLNDEFEDRDAQILGVSIIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQ 120  
Db 61 CPTETIAAFSKLNDEFEDRDAQILGVSIIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQ 120

Qy 121 AAGVLANADGVADRVTFTVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQWNRK 180  
Db 121 AAGVLANADGVADRVTFTVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDDELCAQWNRK 180

Qy 181 GDFPTLDAGELKASA 195  
Db 181 GDFPTLDAGELKASA 195

RESULT 4  
US-10-332-512A-21  
; Sequence 21, Application US/10332512A  
; Publication No. US20040180056A1  
; GENERAL INFORMATION:  
; APPLICANT: ORME, Ian M.  
; APPLICANT: BELISLE, John T.  
; TITLE OF INVENTION: MID-LIFE VACCINE AND METHODS FOR BOOSTING ANTI-MYCOBACTERIAL IMMUN  
; FILE REFERENCE: 38861-186292

```

; CURRENT APPLICATION NUMBER: US/10/332,512A
; CURRENT FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: PCT/US01/21717
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/217,646
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent version 3.1
; SEQ ID NO 21
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-332-512A-21

Query Match
Best Local Similarity 100.0%; Score 1024; DB 4; Length 195;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPLLITGQFPAYQLTALIGDLSKVDAKQPGDYFTTITSDEHPGKRWVFFWPKDFTFV 60
Db 1 MPLLITGQFPAYQLTALIGDLSKVDAKQPGDYFTTITSDEHPGKRWVFFWPKDFTFV 60
Qy 61 CPTETAAFSKLNDEPEDDAQILGVSIIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQ 120
Db 61 CPTETAAFSKLNDEPEDDAQILGVSIIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQ 120
Qy 121 AAGVLANADGVADRVTFIVDPNNEIOFVSATAGSVGRNVDEVLRVLDALQSDLCACMWRK 180
Db 121 AAGVLANADGVADRVTFIVDPNNEIOFVSATAGSVGRNVDEVLRVLDALQSDLCACMWRK 180
Qy 181 GDPTLDAGELLKASA 195
Db 181 GDPTLDAGELLKASA 195

RESULT 5
US-10-732-923-21281
; Sequence 21281, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgettton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 21281
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-10-732-923-21281

Query Match
Best Local Similarity 100.0%; Score 1024; DB 5; Length 195;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPLLITGQFPAYQLTALIGDLSKVDAKQPGDYFTTITSDEHPGKRWVFFWPKDFTFV 60
Db 1 MPLLITGQFPAYQLTALIGDLSKVDAKQPGDYFTTITSDEHPGKRWVFFWPKDFTFV 60
Qy 61 CPTETAAFSKLNDEPEDDAQILGVSIIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQ 120
Db 61 CPTETAAFSKLNDEPEDDAQILGVSIIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQ 120
Qy 121 AAGVLANADGVADRVTFIVDPNNEIOFVSATAGSVGRNVDEVLRVLDALQSDLCACMWRK 180
Db 121 AAGVLANADGVADRVTFIVDPNNEIOFVSATAGSVGRNVDEVLRVLDALQSDLCACMWRK 180
Qy 181 GDPTLDAGELLKASA 195
Db 181 GDPTLDAGELLKASA 195
```

```

RESULT 6
US-10-732-923-21020
; Sequence 21020, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgettton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10,154
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 21020
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Mycobacterium marinum
US-10-732-923-21020

Query Match
Best Local Similarity 92.1%; Score 943; DB 5; Length 195;
Matches 177; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MPLLITGQFPAYQLTALIGDLSKVDAKQPGDYFTTITSDEHPGKRWVFFWPKDFTFV 60
Db 1 MPLLITGQFPAYQLTALIGDLSKVDAKQPGDYFTTITSDEHPGKRWVFFWPKDFTFV 60
Qy 61 CPTETAAFSKLNDEPEDDAQILGVSIIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQ 120
Db 61 CPTETAAFSKLNDEPEDDAQILGVSIIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQ 120
Qy 121 AAGVLANADGVADRVTFIVDPNNEIOFVSATAGSVGRNVDEVLRVLDALQSDLCACMWRK 180
Db 121 AAGVLANADGVADRVTFIVDPNNEIOFVSATAGSVGRNVDEVLRVLDALQSDLCACMWRK 180
Qy 181 GDPTLDAGELLKASA 195
Db 181 GDPTLDAGELLKASA 195

RESULT 7
US-10-282-122A-61886
; Sequence 61886, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Hlangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/220,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
```





```

; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 63991
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-10-282-122A-63991
```

```

Query Match      88.3%; Score 904; DB 4; Length 195;
Best Local Similarity 86.7%; Pred. No. 4.1e-90;
Matches 169; Conservative 12; Mismatches 14; Indels 0; Gaps 0;
```

```

Qy      1 MPLTTIGQPPAYQUTALIGDLSKVDAKOPGDTTITSDHFGKRWVFFWPKDFTFV 60
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1 MSLTISIGQFPAYQUTALIGDLSKVDAKOPGDTTIVSSSHPKRWVFFWPKDFTFI 60
Qy      61 CPTBIAAFSKLUNDEFEDRDAQIIGVSIIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQ 120
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      61 CPTBIAAFSKLUNDEFEGGAQIIGVSIIDSEFHFQWRAQHNDLKTLPFPMLSDIKRELSA 120
Qy      121 AAGVLNADGVADRTFTFVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDCLCACMWRK 180
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      121 ASGALNADGVADRTFTFVDPNDIQFVSATAGSVGRNVEVLRVLDALQSDQLCACMWRK 180
Qy      181 GDPTLDAGELLKASA 195
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      181 GDPTLNATELLKTSA 195
```

```

RESULT 11
US-10-732-923-21017
; Sequence 21017, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; PRIOR FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 21017
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-10-732-923-21017
```

```

Query Match      88.3%; Score 904; DB 5; Length 195;
Best Local Similarity 86.7%; Pred. No. 4.1e-90;
Matches 169; Conservative 12; Mismatches 14; Indels 0; Gaps 0;
```

```

Qy      1 MPLTTIGQPPAYQUTALIGDLSKVDAKOPGDTTITSDHFGKRWVFFWPKDFTFV 60
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      1 MSLTISIGQFPAYQUTALIGDLSKVDAKOPGDTTIVSSSHPKRWVFFWPKDFTFI 60
Qy      61 CPTBIAAFSKLUNDEFEDRDAQIIGVSIIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQ 120
```

```

      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      61 CPTBIAAFSKLUNDEFEGGAQIIGVSIIDSEFHFQWRAQHNDLKTLPFPMLSDIKRELSA 120
Qy      121 AAGVLNADGVADRTFTFVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDCLCACMWRK 180
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      121 ASGALNADGVADRTFTFVDPNDIQFVSATAGSVGRNVEVLRVLDALQSDQLCACMWRK 180
Qy      181 GDPTLDAGELLKASA 195
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      181 GDPTLNATELLKTSA 195
```

```

RESULT 12
US-10-732-923-21019
; Sequence 21019, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; PRIOR FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 21019
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Mycobacterium smegmatis
US-10-732-923-21019
```

```

Query Match      85.7%; Score 878; DB 5; Length 195;
Best Local Similarity 82.2%; Pred. No. 2.8e-87;
Matches 157; Conservative 20; Mismatches 14; Indels 0; Gaps 0;
```

```

Qy      1 MPLTTIGQPPAYQUTALIGDLSKVDAKOPGDTTITSDHFGKRWVFFWPKDFTFV 60
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      1 MALTTIGDQPEYDLTAVVGDSLKVDAKOPDDYFTVTSKDYEGKRWIIFFWPKDFTFV 60
Qy      61 CPTBIAAFSKLUNDEFEDRDAQIIGVSIIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQ 120
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      61 CPTBIAAFSKLUNDEFEDRDAKVLGVSVDNEVHFQWRAQHNDLKTLPFPMVSDIKRELSA 120
Qy      121 AAGVLNADGVADRTFTFVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDCLCACMWRK 180
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      121 ACGVLNADGVADRTFTFVDPNNEIQFVSATAGSVGRNVDEVLRVLDALQSDCLCACMWRK 180
Qy      181 GDPTLDAGELL 191
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      181 GDPTLNAGELL 191
```

```

RESULT 13
US-10-282-122A-53752
; Sequence 53752, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cary, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EITPA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
```



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using SW model

Run on: March 23, 2006, 06:17:23 ; Search time 4.67951 Seconds  
(without alignments)  
1192.746 Million cell updates/sec

Title: US-10-617-038-26  
Sequence: 1 MFLTLTGDPFAYQTLALIG.....CNMRKDPPTLDAGELKASA 195

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 169630 seqs, 28622889 residues

Total number of hits satisfying chosen parameters: 169630

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Published Applications\_AJ\_New\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query	Match length	ID	Description
1	350.5	34.2	446	US-11-108-172-1121	Sequence 1121, Ap
2	318	31.1	206	US-10-821-234-1038	Sequence 1038, Ap
3	318	31.1	257	US-10-642-272A-6	Sequence 6, Appl
4	317	31.0	256	US-10-642-272A-4	Sequence 4, Appl
5	317	31.0	257	US-10-642-272A-5	Sequence 5, Appl
6	312	30.5	198	US-10-642-272A-28	Sequence 28, Appl
7	298	29.1	193	US-10-527-771-6	Sequence 6, Appl
8	274	26.8	187	US-11-058-926-22	Sequence 22, Appl
9	271.5	26.5	271	US-10-821-234-1419	Sequence 1419, Ap
10	263.5	25.7	188	US-11-058-926-24	Sequence 24, Appl
11	261.5	25.5	187	US-11-077-619-4	Sequence 4, Appl
12	259.5	25.3	187	US-11-077-619-86	Sequence 86, Appl
13	210.5	20.6	216	US-11-096-568A-33952	Sequence 33952, A
14	210.5	20.6	229	US-11-096-568A-657	Sequence 657, App
15	210.5	20.6	245	US-11-096-568A-656	Sequence 656, App
16	210.5	20.6	252	US-11-096-568A-33951	Sequence 33951, A
17	142.5	13.9	216	US-11-096-568A-1037	Sequence 1037, Ap
18	142.5	13.9	226	US-11-096-568A-1038	Sequence 1038, Ap
19	140	13.7	180	US-11-096-568A-658	Sequence 658, App
20	133	13.0	218	US-11-096-568A-21440	Sequence 21440, A
21	128	12.5	164	US-11-058-926-19	Sequence 19, Appl
22	103	10.1	32	US-11-058-926-14	Sequence 14, Appl
23	102	10.0	32	US-11-058-926-14	Sequence 14, Appl
24	96.5	9.4	167	US-10-467-657-2342	Sequence 2342, Ap
25	95	9.3	32	US-11-058-926-20	Sequence 20, Appl

26	92.5	9.0	32	7	US-11-058-926-12	Sequence 12, Appl
27	92.5	9.0	32	7	US-11-058-926-13	Sequence 13, Appl
28	92	9.0	164	6	US-10-793-626-252	Sequence 252, App
29	91	8.9	32	7	US-11-058-926-18	Sequence 18, Appl
30	85.5	8.3	32	7	US-11-058-926-17	Sequence 17, Appl
31	83	8.1	164	7	US-11-074-176-70	Sequence 70, Appl
32	82	8.0	376	7	US-11-055-822-422	Sequence 422, App
33	82	8.0	518	7	US-11-055-822-420	Sequence 420, App
34	81	7.9	32	7	US-11-058-926-15	Sequence 15, Appl
35	80	7.8	162	6	US-10-467-657-1176	Sequence 1176, Ap
36	77.5	7.6	168	7	US-11-087-099-10829	Sequence 10829, A
37	77.5	7.6	3361	6	US-10-453-372-1082	Sequence 1082, Ap
38	77	7.5	32	7	US-11-058-926-16	Sequence 16, Appl
39	77	7.5	200	7	US-11-087-099-4145	Sequence 4145, Ap
40	77	7.5	307	7	US-11-096-568A-20009	Sequence 20009, A
41	77	7.5	348	7	US-11-096-568A-20008	Sequence 9, Appl
42	76	7.4	16	7	US-11-058-926-9	Sequence 9, Appl
43	76	7.4	1307	7	US-11-072-512-2438	Sequence 2438, Ap
44	74	7.2	230	7	US-11-087-099-5639	Sequence 5639, Ap
45	74	7.2	231	7	US-11-087-099-5149	Sequence 9149, Ap

## ALIGNMENTS

RESULT 1  
US-11-108-172-1121

Sequence 1121, Application US/11108172

Publication No. US20050260177A1

GENERAL INFORMATION:

APPLICANT: Xu, Jianshun

APPLICANT: Lodes, Michael J.

APPLICANT: Secretist, Heather

APPLICANT: Benson, Darin R.

APPLICANT: Meagher, Madeleine Joy

APPLICANT: Stoik, John A.

APPLICANT: Wang, Tonglong

APPLICANT: Jiang, Yugu

APPLICANT: Smith, Carole L.

APPLICANT: King, Gordon E.

APPLICANT: Wang, Aijun

APPLICANT: Clapper, Jonathan D.

APPLICANT: Skeiky, Yasir A. W.

APPLICANT: Fanger, Gary R.

APPLICANT: Vedvick Thomas S.

APPLICANT: Carter, Darick

TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS

TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE

FILE REFERENCE: 210121.471C15

CURRENT APPLICATION NUMBER: US/11/108, 172

CURRENT FILING DATE: 2005-04-15

PRIOR APPLICATION NUMBER: US 10/025,380

PRIOR FILING DATE: 2001-12-19

PRIOR APPLICATION NUMBER: US 09/922,217

PRIOR FILING DATE: 2001-08-03

PRIOR APPLICATION NUMBER: US 09/833,263

PRIOR FILING DATE: 2001-04-10

PRIOR APPLICATION NUMBER: US 09/649,811

PRIOR FILING DATE: 2000-08-28

PRIOR APPLICATION NUMBER: US 09/609,448

PRIOR FILING DATE: 2000-06-29

PRIOR APPLICATION NUMBER: US 09/575,251

PRIOR FILING DATE: 2000-05-19

PRIOR APPLICATION NUMBER: US 09/519,444

PRIOR FILING DATE: 2000-03-06

PRIOR APPLICATION NUMBER: US 09/504,629

PRIOR FILING DATE: 2000-02-15

PRIOR APPLICATION NUMBER: US 09/480,321

PRIOR FILING DATE: 2000-01-10

PRIOR APPLICATION NUMBER: US 09/476,296

PRIOR FILING DATE: 1999-12-30

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 1130

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1121
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-108-172-1121
```

```
Query Match          34.2%; Score 350.5; DB 7; Length 446;
Best Local Similarity 39.4%; Pred. No. 5,2e-28;
Matches 74; Conservative 35; Mismatches 60; Indels 19; Gaps 4;
```

```
Qy 6 IGDGPAYQLTALIGDLSKVDAKQPGDYFTTTSDEHSGKRVVFPWPKDFTVCPTPEI 65
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 8 INSPAPSEEVVALM-----PNSGFKKISLSYKGMVLLFFPLDFTVCPTPEV 56
Qy 66 AAFSKLNDFFEDRDQAQILGVSIDSEFAHFORAQ---HNDKLTLPFPMLSDIKRELQA 122
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 57 IAFSGSVSRFBNELNEVLAAGSIDSEYALQWLTQDRKKGGIGCTMAIPMLADKTSIASY 116
Qy 123 GVLNADGVADRVTFIVDPNNEIQFVSATAGSVGNVDEVLRVLDALQ---SDELCA 177
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 117 GVLSESGVAVRGLFTIDPHGMLRQITVNDMPVGRSVSEVRLLEAFQFVEKHGEVCPAN 176
Qy 178 WRKGDPTL 185
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 177 WKKGAPTM 184
```

```
RESULT 2
US-10-821-234-1038
; Sequence 1038, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1038
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1038
```

```
Query Match          31.1%; Score 318; DB 6; Length 206;
Best Local Similarity 36.7%; Pred. No. 3,9e-25;
Matches 69; Conservative 33; Mismatches 66; Indels 20; Gaps 4;
```

```
Qy 6 IGDGPAYQLTALIGDLSKVDAKQPGDYFTTTSDEHSGKRVVFPWPKDFTVCPTPEI 65
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 16 IGRAPAPFKATVAVD-----AFKEVKLSIDYKGVLLFFPLDFTVCPTPEI 63
Qy 66 AAFSKLNDFFEDRDQAQILGVSIDSEFAHFOR---RAQHNLDKLPFPMLSDIKRELQA 122
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 64 IAFSNRAEDFRKGLGCEVIGSVDSQFTHLAMINTPRKEGGAGPINTIPLDLVTRLSBDY 123
Qy 123 GVLNADGVADRVTFIVDPNNEIQFVSATAGSVGNVDEVLRVLDALQ---SDELCA 177
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 124 GVLKTDGSIARGLFTIDGKVLKQITVNDLPVGRSVDELRLVQAFQYTDHGEVCPAG 183
Qy 178 WRKGDPTL 185
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 184 WKPGSDTI 191
```

```
RESULT 3
```

```
US-10-642-272A-6
; Sequence 6, Application US/10642272A
; Publication No. US20050277606A1
; GENERAL INFORMATION:
; APPLICANT: Hattori, Fumiyuki
; APPLICANT: Sugimura, Kei-iro
; APPLICANT: Furuya, Mayumi
; TITLE OF INVENTION: Therapeutic Methods and Agents for Diseases Associated with
; FILE REFERENCE: 58777.000012
; CURRENT APPLICATION NUMBER: US/10/642,272A
; PRIOR APPLICATION NUMBER: PCT/JP02/01358
; PRIOR FILING DATE: 2001-02-18
; PRIOR APPLICATION NUMBER: JP 41003/2001
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Version 3.3
; SEQ ID NO 6
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Mus sp.
US-10-642-272A-6
```

```
Query Match          31.1%; Score 318; DB 6; Length 257;
Best Local Similarity 37.2%; Pred. No. 5,3e-25;
Matches 68; Conservative 38; Mismatches 57; Indels 20; Gaps 4;
```

```
Qy 11 PAYQLTALIGDLSKVDAKQPGDYFTTTSDEHSGKRVVFPWPKDFTVCPTPEIAFSK 70
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 71 PYFKGIVAVNGE-----FKELSDDRKGLVLFYPLDFTVCPTPEIYAFSD 118
Qy 71 LNDEFEDRDQAQILGVSIDSEFAHFOR---RAQHNLDKLTLPFPMLSDIKRELQAAGV-LN 126
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 119 KANEFHDVNCVAVASVDSHFSHLAMINTPRKKGAGHNMITLSDITKQISHDYGLLE 178
Qy 127 ADGVADRVTFIVDPNNEIQFVSATAGSVGNVDEVLRVLDALQ---SDELCA 182
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 179 SAGIALRGFLTIDPNGVKHLVNDLPVGRSVETRLVAVAFQFVETHEGVCANWTPES 238
Qy 183 PTL 185
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 239 PTI 241
```

```
RESULT 4
US-10-642-272A-4
; Sequence 4, Application US/10642272A
; Publication No. US20050277606A1
; GENERAL INFORMATION:
; APPLICANT: Hattori, Fumiyuki
; APPLICANT: Sugimura, Kei-iro
; APPLICANT: Furuya, Mayumi
; TITLE OF INVENTION: Therapeutic Methods and Agents for Diseases Associated with
; FILE REFERENCE: 58777.000012
; CURRENT APPLICATION NUMBER: US/10/642,272A
; PRIOR APPLICATION NUMBER: PCT/JP02/01358
; PRIOR FILING DATE: 2001-02-18
; PRIOR APPLICATION NUMBER: JP 41003/2001
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Version 3.3
; SEQ ID NO 4
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-642-272A-4
```

```
Query Match          31.0%; Score 317; DB 6; Length 256;
Best Local Similarity 37.2%; Pred. No. 6,6e-25;
Matches 68; Conservative 37; Mismatches 58; Indels 20; Gaps 4;
```

```

QY 11 PAYQLTALIGDLSKVDAKQPGDYFTTTITSDHEPGKRRVVFPPKQFTVCPPEIAAFSK 70
Db 70 PYFKTIAVNGE-----FKDLSLDPFKGYLVLFYBPDLFTVCPPEIAVAFSD 117
QY 71 LNDEFEDRDAQILGVSIDSEFAHPOW---RAQINDKLTLPPLMSDIKRELSOAGV-LN 126
Db 118 KANEHFDVANCEVYANSVDSHFSLHMINIPRKKGIGJGHNNIALISDITKQISXDYVLE 177
QY 127 ADGVADRVTFIYDPNNIEIQFVSATAGSVGRNDEVLRVLDALQ-----SDELCAKNRKGD 182
Db 178 GSGIALRLRGFLIDPNGVIKHLSVNDLPGVRSVETRLVYKAFQYVETHGEVCPANWTPDS 237
QY 183 PTL 185
Db 238 PTL 240

```

```

RESULT 5
US-10-642-272A-5
; Sequence 5, Application US/10642272A
; Publication No. US20050277606A1
; GENERAL INFORMATION:
; APPLICANT: Hattori, Fumiyuki
; APPLICANT: Sugimura, Keijiro
; APPLICANT: Futuya, Mayumi
; TITLE OF INVENTION: Therapeutic Methods and Agents for Diseases Associated with
; TITLE OF INVENTION: Decreased Expression of AOP-1 Gene or AOP-1
; FILE REFERENCE: 58777.000012
; CURRENT APPLICATION NUMBER: US/10/642, 272A
; PRIORITY FILING DATE: 2003-08-18
; PRIOR APPLICATION NUMBER: PCT/JP02/01558
; PRIORITY FILING DATE: 2001-02-18
; PRIOR APPLICATION NUMBER: JP 41003/2001
; PRIORITY FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 5
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-642-272A-5

```

[illegible]

RESULT 6  
US-10-642-272A-28  
; Sequence 28, Application US/10642272A  
; Publication No. US20050277606A1  
; GENERAL INFORMATION:  
; APPLICANT: Hatoori, FumiYuki  
; APPLICANT: Sugimura, Keijiro  
; APPLICANT: Furiya, Mayumi  
; TITLE OF INVENTION: Therapeutic Methods and Agents for Diseases Associated with

```

/ TITLE OF INVENTION: Decreased Expression of AOP-1 Gene or AOP-1
/ FILE REFERENCE: 58777.000012
/ CURRENT APPLICATION NUMBER: US/10/642,272A
/ CURRENT FILING DATE: 2003-08-18
/ PRIOR APPLICATION NUMBER: PCT/JPO02/01358
/ PRIOR FILING DATE: 2001-02-18
/ PRIOR APPLICATION NUMBER: JP 41003/2001
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 28
/ LENGTH: 198
/ TYPE: PRT
/ ORGANISM: Rattus norvegicus
US-10-642-272A-28

```

Query Match	30.5%	Score 312;	DB 6;	Length 198;
Best Local Similarity	37.2%	Pred. No. 1.5e-24;		
Matches	70;	Conservative 32;	Mismatches 66;	Indels 20; Gaps 4;

  

QY	6	IGDGFAYQLTALLIGDLISKYDAKAPGDYFTTITSDEHGMKRVVFWPKDFTVCPTETI	65
DB	8	IGKPAPEPTGAVVDG-----AFKEIKLSDYRGKVVLLFFPYLDFTFCPTETI	55
QY	66	AAFTSLNDEFEEDRAQILGVSIIDSEFAHFW---RAGHNDLKTLPFLPLSDIKELSGAA	122
DB	56	IATSPHADDFKLCCEVLGVSVDSQFTFLAIINTPRKEGSGIPNTIPLLADVTSLSQNY	115
QY	123	GVLT-NADGVARVTFIVDPNNIEIQFVSATAGSVGRNDEVLRVLDALQ---SDELCACN	177
DB	116	GVLNKNDSEIAIRGFIIDAKGVNLGITVNDLPVGRSVDEALRTVQAQFYTDHEHGVCPAG	175
QY	178	WRKEDPTL 185	
DB	176	WKPGSDTI 183	

```

RESULT 7
US-10-527-771-6
; Sequence 6, Application US/10527771
; Publication No. US20050271683A1
; GENERAL INFORMATION:
; APPLICANT: University Gent
; TITLE OF INVENTION: Ostertagia vaccine
; FILE REFERENCE: 2002-015
; CURRENT APPLICATION NUMBER: US/10/527,771
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: US 10/243,319
; PRIOR FILING DATE: 2002-09-13
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Ostertagia ostertagi
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: (72)..(72)
; OTHER INFORMATION: The xaa at location 72 stands for Asn, or Thr.
US-10-527-771-6

```

	Query Match	29.1%	Score	298	DB	6	Length	193
	Best Local Similarity	35.6%	Pred.	No.	4e-23			
	Matches	68	Conservative	31	Mismatches	72	Indels	20
					Gaps	4		
QY	6	IGDGPAYQLTALIGGDISKVDAKOPGDYFTTTISDEHPGRKRVVFPFKDFTCVCPREI	65					
DB	3	IGKPAPDAITAAVYNGD-----FDIVKLSDYDKGYRTVLPFLPDLFTVCPEI	50					
QY	66	AAFSKLNDDEFDRDAQILGVSIDSEFAHFOW--RAQHNDIKTLVPFPLMSIDIKELSGAA	122					
DB	51	IAPSPRVEEFPKKIIAVALACXDSVFSLHAIINTPRKKGSGAGDNMIIPVLADNTNIOADXY	110					

Qy 123 GVADRVTFIVDPNNELQFVSATAGSVGRNDEVLRVLDALQ----SDELCAKN 177  
Db 111 GVLKEDDEIAVRGGLFIDPKGILRQITVNDLPVGRSVDETLRLVQAFQYVDKRGVCPAG 170  
Qy 178 WRKGDPTLDAG 188  
Db 171 WTPGKATIKPG 181

## RESULT 8

US-11-058-926-22  
; Sequence 22, Application US/11058926  
; Publication No. US20060030022A1  
; GENERAL INFORMATION:  
; APPLICANT: BECKWITH, JONATHAN  
; APPLICANT: ASLUND, FREDRIK  
; APPLICANT: BESSETTE, PAUL H.  
; APPLICANT: GEORGIU, GEORGE  
; APPLICANT: RITZ, DANIEL  
; APPLICANT: LIM, JACKIE EUN-AH  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PRODUCTION OF DISULFIDE  
; FILE REFERENCE: HMV-052.01  
; CURRENT APPLICATION NUMBER: US/11/058,926  
; CURRENT FILING DATE: 2005-02-16  
; PRIOR APPLICATION NUMBER: US/09/679,705  
; PRIOR FILING DATE: 2000-10-05  
; PRIOR APPLICATION NUMBER: 60/157,770  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: 60/163,939  
; PRIOR FILING DATE: 1999-11-08  
; PRIOR APPLICATION NUMBER: 60/166,044  
; PRIOR FILING DATE: 1999-11-17  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patentl Ver. 2.1  
; SEQ ID NO 22  
; LENGTH: 187  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-11-058-926-22

Query Match 26.8%; Score 274; DB 7; Length 187;  
Best Local Similarity 34.4%; Pred. No. 1.1e-20;  
Matches 54; Conservative 29; Mismatches 68; Indels 6; Gaps 2;  
Qy 35 FTTTSDHPRGKRVVFEPMDFTVCPTETLAASKLNDEFEDRDAQILGVSIDSEPAHF 94  
Db 21 FIETKDTGRKSVFFFPADFTVCPTETLGDVADHDELQKLGVDVYASTDTHFTK 80  
Qy 95 QWRAGHNDLKTLPFPMLSIDIRELSQAAGVLNAD-GVADRVTFIVDPNNELQFVSATAGS 153  
Db 81 AMHSSETIAKIKYAMIGDPTGALTFRNFDNMRDEGLADRAATFVVDPOGIIQALEVTAEG 140  
Qy 154 VGRVNDVLRVLDALQ-----SDELCAKNRKGDPTL 185  
Db 141 IGRDASDLRKIKAAQYVASHPGVCAPKMKGEATL 177

## RESULT 9

US-10-821-234-1419  
; Sequence 1419, Application US/10821234  
; Publication No. US2005025514A1  
; GENERAL INFORMATION:  
; APPLICANT: Labat, Ivan  
; APPLICANT: Stache-Crain, Blythe  
; APPLICANT: Andaman, Susan  
; APPLICANT: Tang, Y. Tom  
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia  
; FILE REFERENCE: 821A  
; CURRENT APPLICATION NUMBER: US/10/821,234  
; CURRENT FILING DATE: 2004-04-07  
; PRIOR APPLICATION NUMBER: US 60/462,047  
; PRIOR FILING DATE: 2003-04-07

; NUMBER OF SEQ ID NOS: 1704  
; SOFTWARE: PC\_SEQ\_genes Version 1.0  
; SEQ ID NO 1419  
; LENGTH: 271  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-821-234-1419

Query Match 26.5%; Score 271.5; DB 6; Length 271;  
Best Local Similarity 33.9%; Pred. No. 3.2e-20;  
Matches 64; Conservative 32; Mismatches 72; Indels 21; Gaps 5;

Qy 23 LSKIDAKQPGDY-----FTTSDHPRGKRVVFEPMDFTVCPTETLAASKLN 72  
Db 76 LSKAKISKPAWEGTAVIDGEPKELTTRYRGYLVFFYPPLDFTVCPTETLAASKLN 135  
Qy 73 DEFEDRDAQILGVSIDSEPAHFQW--RAQHNDLKTLPFPMLSIDIRELSQAAGV-LNAD 128  
Db 136 EEFKSTIEVYAGSVDSQFTHLAMINTPRQGLGPRIRILSDTHQIKDYGVLEDS 195  
Qy 129 GVADRVTFIVDPNNELQFVSATAGSVGRNDEVLRVLDALQ---SDELCAKNRKGDPT 184  
Db 196 GHTLRGLFIIIDKGLRQITVNDLPVGRSVDETLRLVQAFQYVDKRGVCPAGKPGSET 255  
Qy 185 L--DAGEL 190  
Db 256 ITPDPAGKL 264

## RESULT 10

US-11-058-926-24  
; Sequence 24, Application US/11058926  
; Publication No. US20060030022A1  
; GENERAL INFORMATION:  
; APPLICANT: BECKWITH, JONATHAN  
; APPLICANT: ASLUND, FREDRIK  
; APPLICANT: BESSETTE, PAUL H.  
; APPLICANT: GEORGIU, GEORGE  
; APPLICANT: RITZ, DANIEL  
; APPLICANT: LIM, JACKIE EUN-AH  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PRODUCTION OF DISULFIDE  
; FILE REFERENCE: HMV-052.01  
; CURRENT APPLICATION NUMBER: US/11/058,926  
; CURRENT FILING DATE: 2005-02-16  
; PRIOR APPLICATION NUMBER: US/09/679,705  
; PRIOR FILING DATE: 2000-10-05  
; PRIOR APPLICATION NUMBER: 60/157,770  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: 60/163,939  
; PRIOR FILING DATE: 1999-11-08  
; PRIOR APPLICATION NUMBER: 60/166,044  
; PRIOR FILING DATE: 1999-11-17  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patentl Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 188  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-11-058-926-24

Query Match 25.7%; Score 263.5; DB 7; Length 188;  
Best Local Similarity 34.2%; Pred. No. 1.3e-19;  
Matches 54; Conservative 29; Mismatches 68; Indels 7; Gaps 3;

Qy 35 FTTTSDHPRGKRV-VFEPMDFTVCPTETLAASKLNDEFEDRDAQILGVSIDSEPAH 93  
Db 21 FIETKDTGRKSVFFFPADFTVCPTETLGDVADHDELQKLGVDVYASTDTHFTK 80  
Qy 94 QWRAGHNDLKTLPFPMLSIDIRELSQAAGVLNAD-GVADRVTFIVDPNNELQFVSATAG 152  
Db 81 KAHSSSETIAKIKYAMIGDPTGALTFRNFDNMRDEGLADRAATFVVDPOGIIQALEVTA 140

Qy 153 SVGRNDEVLRVLDALQ-----SDELCACNMRKGDPTL 185  
Db 141 GIGRDASDLKRKIKAAQYVASHPGSEVCPAKWEGEATL 178

## RESULT 11

US-11-077-619-4  
Sequence 4, Application US/11077619  
Publication No. US20060040279A1  
GENERAL INFORMATION:  
APPLICANT: Feesche, Joerg  
APPLICANT: Maurer, Karl-Heinz  
APPLICANT: Breves, Roland  
APPLICANT: Schneider, Thomas  
APPLICANT: Hecker, Michael  
APPLICANT: Juergen, Britta  
APPLICANT: Voigt, Birgit  
TITLE OF INVENTION: DNA CHIPS USED FOR BIOPROCESS CONTROL  
FILE REFERENCE: HENK-0122 / H5692  
CURRENT APPLICATION NUMBER: US/11/077,619  
PRIOR FILING DATE: 2005-03-11  
PRIOR APPLICATION NUMBER: PCT/EP2003/009979  
PRIOR FILING DATE: 2003-09-09  
PRIOR APPLICATION NUMBER: DE 10242433.0  
NUMBER OF SEQ ID NOS: 130  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 4  
LENGTH: 187  
TYPE: PRT  
ORGANISM: Bacillus subtilis  
US-11-077-619-4

Query Match 25.5%; Score 261.5; DB 7; Length 187;  
Best Local Similarity 31.6%; Pred. No. 2.1e-19;  
Matches 56; Conservative 39; Mismatches 73; Indels 9; Gaps 4;

Qy 17 ALIGDLSKVDAK--QPGDYFTTTSDEHPGKRVVFWPMDFTVCPTETIAAFSKLNDE 74  
Db 2 SLIGKEVPPFAKAKYKNGE-FIDVTNEHLKGQMSVFCYFPADFSVCTELEDLOEQYAA 60  
Qy 75 FEDRDAQILGVSIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQAAGVLAN-ADGVA 133  
Db 61 LKEIGVEVYSVSTDTHTVHKGWHDSSSEKISKITYAMIGDSPQTSIRNPDVLDDEETGLDR 120  
Qy 134 VTFTVPDNNETQFVSATGSGVGRNDEVLRVLDALQ-----SDELCACNMRKGDPTL 185  
Db 121 GTFTIDPDGVLTQAVEINAGGIGRDASNLVNKKAQYVRONPGSEVCPAKWEGEATL 177

## RESULT 12

US-11-077-619-86  
Sequence 86, Application US/11077619  
Publication No. US20060040279A1  
GENERAL INFORMATION:  
APPLICANT: Feesche, Joerg  
APPLICANT: Maurer, Karl-Heinz  
APPLICANT: Breves, Roland  
APPLICANT: Schneider, Thomas  
APPLICANT: Hecker, Michael  
APPLICANT: Juergen, Britta  
APPLICANT: Voigt, Birgit  
TITLE OF INVENTION: DNA CHIPS USED FOR BIOPROCESS CONTROL  
FILE REFERENCE: HENK-0122 / H5692  
CURRENT APPLICATION NUMBER: US/11/077,619  
PRIOR FILING DATE: 2005-03-11  
PRIOR APPLICATION NUMBER: PCT/EP2003/009979  
PRIOR FILING DATE: 2003-09-09  
PRIOR APPLICATION NUMBER: DE 10242433.0  
NUMBER OF SEQ ID NOS: 130  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 86

LENGTH: 187  
TYPE: PRT  
ORGANISM: Bacillus licheniformis  
US-11-077-619-86

Query Match 25.3%; Score 259.5; DB 7; Length 187;  
Best Local Similarity 31.6%; Pred. No. 3.3e-19;  
Matches 56; Conservative 37; Mismatches 75; Indels 9; Gaps 4;

Qy 17 ALIGDLSKVDAK--QPGDYFTTTSDEHPGKRVVFWPMDFTVCPTETIAAFSKLNDE 74  
Db 2 SLIGKEVPPFAKAKYKNGE-FIDVTNEHLKGQMSVFCYFPADFSVCTELEDLOEQYAA 60  
Qy 75 FEDRDAQILGVSIDSEFAHFQWRAQHNDLKTLPFPMLSDIKRELSQAAGVLAN-ADGVA 133  
Db 61 LKEIGVEVYSVSTDTHTVHKGWHDSSSEKISKITYAMIGDSPQTSIRNPDVLDDEETGLDR 120  
Qy 134 VTFTVPDNNETQFVSATGSGVGRNDEVLRVLDALQ-----SDELCACNMRKGDPTL 185  
Db 121 GTFTIDPDGVLTQAVEINAGGIGRDASILVNKKAQYVRONPGSEVCPAKWEGEATL 177

## RESULT 13

US-11-096-568A-33952  
Sequence 33952, Application US/11096568A  
Publication No. US20060048240A1  
GENERAL INFORMATION:  
APPLICANT: Alexandrov, Nickolai et al.  
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
FILE REFERENCE: 2750-1592PUS2  
CURRENT APPLICATION NUMBER: US/11/096,568A  
CURRENT FILING DATE: 2005-04-01  
NUMBER OF SEQ ID NOS: 34471  
SEQ ID NO 33952  
LENGTH: 216  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
FEATURES:  
NAME/KEY: misc feature  
LOCATION: (1)..(216)  
OTHER INFORMATION: Cerec Seq. ID no. 13604610  
US-11-096-568A-33952

Query Match 20.6%; Score 210.5; DB 7; Length 216;  
Best Local Similarity 28.2%; Pred. No. 4.1e-14;  
Matches 55; Conservative 38; Mismatches 77; Indels 25; Gaps 6;

Qy 1 MPLTTIGDQFPAYQLTALIGDLSKVDAKQPGDYFTTTSDEHPGKRVVFWPMDFTPV 60  
Db 1 MPGLTLDGTVPNLEVE-----TTHDKFKLHDVFA-----NSMTVLFSHSGDFTPV 45  
Qy 61 CPTETIAAFSKLNDEFEEDRDAQILGVSIDSEFAHFQWRAQ--HNDLKTLPFPMLSDIKRE 117  
Db 46 CTTETLGAMAKYAHFDRKGVLTGLSCDVOQSHKDWKIDLEAFHSGSKVVPPIIADBNKE 105  
Qy 118 LSQAAGVLAN-ADGVAADVTFITVPDNNETQFVSATGSGVGRNDEVLRVLDALQ-----QSD 171  
Db 106 IIPQLANMIDPLENGSPALHIVGPDSSKIKSLFYSTTGKRWDEVLRALDLSMASTKNN 165  
Qy 172 ELCA-CNMRKGDPTL 185  
Db 166 KIATPVMMKPDQPVV 180

## RESULT 14

US-11-096-568A-657  
Sequence 657, Application US/11096568A  
Publication No. US20060048240A1  
GENERAL INFORMATION:  
APPLICANT: Alexandrov, Nickolai et al.  
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
FILE REFERENCE: 2750-1592PUS2  
CURRENT APPLICATION NUMBER: US/11/096,568A  
CURRENT FILING DATE: 2005-04-01  
NUMBER OF SEQ ID NOS: 34471  
SEQ ID NO 657



FILE REFERENCE: 2750-1592PUS2  
CURRENT APPLICATION NUMBER: US/11/096,568A  
CURRENT FILING DATE: 2005-04-01  
NUMBER OF SEQ ID NOS: 34471  
SEQ ID NO 657  
LENGTH: 229  
TYPE: PRT  
ORGANISM: Zea mays subsp. mays  
FEATURE:  
NAME/KEY: mlec feature  
LOCATION: (1)..(229)  
OTHER INFORMATION: Cereals Seq. ID no. 12637707  
US-11-096-568A-657

Query Match 20.6%; Score 210.5; DB 7; Length 229;  
Best Local Similarity 27.5%; Pred. No. 4,5e-14;  
Matches 57; Conservative 34; Mismatches 75; Indels 41; Gaps 5;

QY 1 MPGLTIGDPPAYQTLALIGDLSKVDKAKQPDYFTTTSDBHPGK-WRVVFEPKQFTF 59  
DB 1 MPGLTIGDTPVNLBIDSTHG-----KIRIDYVGDGYAIIFSHPADFTF 44  
QY 60 VCPTEIAFSLKUNDEFEDRDQILGVSIIDSEFAHFW-----RAQHNDLKTLPF 108  
DB 45 VCTTEMAAMAGYAKFEKRGVYKLGISCDVESHQMTXDVEAYGKQOQOQATTTKVTF 104  
QY 109 PMLSDIKRELQAAGVLANAD-----GVADRVTFIVDPNNEIOFVSATAGSVGRNVDEV 161  
DB 105 PTLADPARDAIRQLNMDVPDEKDAAGRMSRRLHVVGPDKAVKLSFLYPATTTGRNDEV 164  
QY 162 LRVLDALQS-----DELCAQNRKGD 182  
DB 165 LRAVDSLITLAAKHGKGVATPAMWKPE 191

RESULT 15  
US-11-096-568A-656  
Sequence 656, Application US/11096568A  
Publication No. US20060048240A1  
GENERAL INFORMATION:  
APPLICANT: Alexandrov, Nickolai et al.  
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
FILE REFERENCE: 2750-1592PUS2  
CURRENT APPLICATION NUMBER: US/11/096,568A  
CURRENT FILING DATE: 2005-04-01  
NUMBER OF SEQ ID NOS: 34471  
SEQ ID NO 656  
LENGTH: 245  
TYPE: PRT  
ORGANISM: Zea mays subsp. mays  
FEATURE:  
NAME/KEY: mlec feature  
LOCATION: (1)..(245)  
OTHER INFORMATION: Cereals Seq. ID no. 12637706  
US-11-096-568A-656

Query Match 20.6%; Score 210.5; DB 7; Length 245;  
Best Local Similarity 27.5%; Pred. No. 4,9e-14;  
Matches 57; Conservative 34; Mismatches 75; Indels 41; Gaps 5;

QY 1 MPGLTIGDPPAYQTLALIGDLSKVDKAKQPDYFTTTSDBHPGK-WRVVFEPKQFTF 59  
DB 17 MPGLTIGDTPVNLBIDSTHG-----KIRIDYVGDGYAIIFSHPADFTF 60  
QY 60 VCPTEIAFSLKUNDEFEDRDQILGVSIIDSEFAHFW-----RAQHNDLKTLPF 108  
DB 61 VCTTEMAAMAGYAKFEKRGVYKLGISCDVESHQMTXDVEAYGKQOQOQATTTKVTF 120  
QY 109 PMLSDIKRELQAAGVLANAD-----GVADRVTFIVDPNNEIOFVSATAGSVGRNVDEV 161  
DB 121 PTLADPARDAIRQLNMDVPDEKDAAGRMSRRLHVVGPDKAVKLSFLYPATTTGRNDEV 180

QY 162 LRVLDALQS-----DELCAQNRKGD 182  
DB 181 LRAVDSLITLAAKHGKGVATPAMWKPE 207

Search completed: March 23, 2006, 06:29:37  
Job time : 5.67951 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 04:56:29 ; Search time 72.3158 Seconds  
(without alignments)  
1652.628 Million cell updates/sec

Title: US-10-617-038-27

Perfect score: 1397  
Sequence: 1 MSGGEPMTKTIIVGIDGSR.....GPSGHFVRHAECVLYVRR 272

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_21:\*

1:	geneeqp1980s:*
2:	geneeqp1980s:*
3:	geneeqp2000s:*
4:	geneeqp2001s:*
5:	geneeqp2002s:*
6:	geneeqp2003as:*
7:	geneeqp2003bs:*
8:	geneeqp2004s:*
9:	geneeqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	1397	100.0	272	8	AD137306 M. tuberc
2	416	29.8	279	8	AD137300 M. tuberc
3	357.5	25.6	268	8	AD137321 M. tuberc
4	286	20.5	295	8	AD137298 M. tuberc
5	267	19.1	297	5	AY87878 M. tuberc
6	267	19.1	297	5	ABG30907 Mycobacte
7	266	19.0	301	4	AB79640 Corynebac
8	266	19.0	301	4	AAU71872 C. glutam
9	266	19.0	301	4	AAU71865 C. glutam
10	266	19.0	301	4	AAU71865 C. glutam
11	266	19.0	301	5	ABG80324 C. glutam
12	266	19.0	301	9	ABE15399 C. glutam
13	249	17.8	300	6	ABU25937 Protein e
14	151	10.8	316	4	AAU62080 Propionib
15	151	10.8	316	6	ABM58599 Propionib
16	128.5	9.2	173	8	ADU05947 Plant ful
17	127	9.1	179	6	ADA48734 Rice prot
18	127	9.1	179	7	ABM90364 Rice prot
19	124	8.9	165	6	ADA48322 Rice prot
20	124	8.9	165	7	ABM86830 Rice abio
21	124	8.9	165	9	ADY60949 Abiotic s
22	121.5	8.7	524	2	AAW34202 Streptomy
23	121.5	8.7	524	2	AAW55803 Streptomy
24	116.5	8.3	162	3	AAQ42273 Arabidops

25	116.5	8.3	198	3	AAQ42272 Arabidops
26	113	8.1	153	3	AAQ42274 Arabidops
27	112.5	8.1	162	3	AAQ54317 Arabidops
28	112.5	8.1	183	3	AAQ54316 Arabidops
29	112	8.0	189	5	ABP38950 Staphyloc
30	112	8.0	189	8	ADSO6786 Staphyloc
31	110	7.9	149	7	ADM25915 Hyperther
32	109.5	7.8	166	6	ABM73534 Staphyloc
33	109	7.8	153	3	AAQ54318 Arabidops
34	109	7.8	158	7	ADM25536 Hyperther
35	109	7.8	292	5	ABAB1331 Human ORF
36	109	7.8	292	5	ABP07225 Human ORF
37	109	7.8	941	6	ABU36688 Protein e
38	109	7.8	941	6	ABU34503 Protein e
39	108	7.7	147	4	AAQ91259 C. glutam
40	108	7.7	245	8	ADY06981 Plant ful
41	107.5	7.7	6751	8	ADQ91702 Polyketid
42	107.5	7.7	6751	8	ABE86594 Streptomy
43	107.5	7.7	6751	8	ABE86791 Streptomy
44	107	7.7	206	8	ADX77663 Plant ful
45	105.5	7.6	613	4	AAU04887 Micromono

## ALIGNMENTS

RESULT 1					
ID	AD137306	standard; protein; 272 AA.			
XX	AD137306;				
AC	AD137306;				
XX	AD137306;				
DT	22-APR-2004	(first entry)			
XX	22-APR-2004				
DE	M. tuberculosis low oxygen induced antigen Rv2624c SEQ ID NO:27.				
KM	mycobacterial infection; vaccine; tuberculosis;				
KW	Mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;				
XX	low oxygen induced antigen.				
OS	Mycobacterium tuberculosis.				
XX					
PN	W02004006952-A2.				
XX					
PD	22-JAN-2004.				
XX					
PF	08-JUL-2003; 2003WO-DK000477.				
XX					
PR	13-JUL-2002; 2002DK-00001098.				
XX					
PA	(STAT-) STATENS SERUM INST.				
XX					
PI	Andersen P, Rosenkrands I, Stryhn A;				
XX					
DR	WPI, 2004-122778/12.				
XX					
PT	N-PSDB; AD137351.				
XX					
PS	Claim 3; SEQ ID NO 27; 76pp; English.				
XX					
CC	The present invention describes polypeptides or their fragments, which				
CC	are expressed during the latent stage of a mycobacterial infection,				
CC	and/or nucleic acids encoding the polypeptides, which are useful for				
CC	creating a therapeutic vaccine against tuberculosis. Also described: (1)				
CC	a therapeutic vaccine against tuberculosis comprising one or more				
CC	polypeptides; (2) a method for treating an animal, including a human				
CC	being, with tuberculosis caused by virulent mycobacteria, e.g. by				
CC	Mycobacterium tuberculosis, M. africanum or M. bovis; (3) a method for				
CC	immunising an animal, including a human being, against tuberculosis				

CC caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis  
CC caused by virulent mycobacteria in an animal, including a human being;  
CC (5) a method for diagnosing previous or ongoing infection with a virulent  
CC mycobacterium; and (6) a method of diagnosing Mycobacterium tuberculosis  
CC infection in a subject. The polypeptides have antibacterial activities,  
CC and can be used in vaccines and in gene therapy. The polypeptides are  
CC useful for the manufacture of a therapeutic vaccine for treating an  
CC individual who is infected by a virulent mycobacterium, e.g. M.  
CC tuberculosis, and who is not vaccinated with BCG against tuberculosis.  
CC The present sequence represents a low oxygen induced antigen, which is  
CC used in the exemplification of the present invention.

SQ Sequence 272 AA;

Query Match 100.0%; Score 1397; DB 8; Length 272;

Best Local Similarity 100.0%; Pred. No. 1.1e-136;

Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSGGEPMTKTIYIGIDSHAITAALMGVDEAISRAVPLVLVYIKETHSPDDYRDL 60  
DB 1 MSGGEPMTKTIYIGIDSHAITAALMGVDEAISRAVPLVLVYIKETHSPDDYRDL 60  
QY AHAERSLREASQSAVEAAKLVKIEITDIPRGAPVLYEASRDAMI CVSGVIGRVAASI 120  
DB AHAERSLREASQSAVEAAKLVKIEITDIPRGAPVLYEASRDAMI CVSGVIGRVAASI 120  
QY 121 LGSTATLEAKACPVAVMRSKVDQPSADINMIVVMTADPDNEAVLEAAREAKLRQAP 180  
DB 121 LGSTATLEAKACPVAVMRSKVDQPSADINMIVVMTADPDNEAVLEAAREAKLRQAP 180  
QY 181 ILAAGRREELREIPDGEFERRVQDMHRRHPDVRYPTTHTGTARFLADHDERVQLAVI 240  
DB 181 ILAAGRREELREIPDGEFERRVQDMHRRHPDVRYPTTHTGTARFLADHDERVQLAVI 240  
QY 241 GGGGAGOLARLVGSPGHPFRHAECVLYVVR 272  
DB 241 GGGGAGOLARLVGSPGHPFRHAECVLYVVR 272

RESULT 2

AD137300 standard; protein; 279 AA.

AC AD137300;

DT 22-APR-2004 (first entry)

DE M. tuberculosis low oxygen induced antigen Rv2028c SEQ ID NO:21.

KM mycobacterial infection; vaccine; tuberculosis;

KM Mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;

XX low oxygen induced antigen.

XX Mycobacterium tuberculosis.

XX WO2004006952-A2.

XX 22-JAN-2004.

XX 08-JUL-2003; 2003WO-DK000477.

XX 13-JUL-2002; 2002DK-00001098.

XX (STAT-) STATENS SERUM INST.

XX Andersen P, Rosenkrands I, Stryhn A;

XX WPI; 2004-122778/12.

XX DR N-PSDB; AD137345.

PT Use of one or more polypeptides or their fragments, which are expressed  
PT during the latent stage of the mycobacterial infection, and/or nucleic  
PT acids encoding the polypeptides, for a therapeutic vaccine against

PT tuberculosis.

XX Claim 3; SEQ ID NO 21; 76pp; English.

XX The present invention describes polypeptides or their fragments, which  
XX are expressed during the latent stage of a mycobacterial infection,  
XX and/or nucleic acids encoding the polypeptides, which are useful for  
XX creating a therapeutic vaccine against tuberculosis. Also described: (1)  
XX a therapeutic vaccine against tuberculosis comprising one or more  
XX polypeptides; (2) a method for treating an animal, including a human  
XX being, with tuberculosis caused by virulent Mycobacterium tuberculosis  
XX Mycobacterium tuberculosis, M. africanum or M. bovis; (3) a method for  
XX immunising an animal, including a human being, against tuberculosis  
XX caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis  
XX caused by virulent mycobacteria in an animal, including a human being;  
XX (5) a method for diagnosing previous or ongoing infection with a virulent  
XX mycobacterium; and (6) a method of diagnosing Mycobacterium tuberculosis  
XX infection in a subject. The polypeptides have antibacterial activities,  
XX and can be used in vaccines and in gene therapy. The polypeptides are  
XX useful for the manufacture of a therapeutic vaccine for treating an  
XX individual who is infected by a virulent mycobacterium, e.g. M.  
XX tuberculosis, and who is not vaccinated with BCG against tuberculosis.  
XX The present sequence represents a low oxygen induced antigen, which is  
XX used in the exemplification of the present invention.

SQ Sequence 279 AA;

Query Match 29.8%; Score 416; DB 8; Length 279;

Best Local Similarity 36.8%; Pred. No. 2.2e-34;

Matches 98; Conservative 46; Mismatches 116; Indels 6; Gaps 2;

QY 11 TIIVGIDGSHAITAALMGVDEAISRAVPLVLVYIKETHSPDDYD--RDIAHAERSL 67  
DB 9 SIIVGIDGSKPAVOALLMAVDEAASRDIPRLLYALEPDDPGYAAAGAAARKLAEMNAV 68  
QY 68 REAQSAYEAGKLVKIEITDIPRGAPVLYEASRDAMI CVSGVIGRVAASITLGSTATE 127  
DB 69 RYAFYAEADRPVKVEITQERPVTSILIRASAAALVCVGAIVGHHFPERVGTAA 128  
QY 128 LAEKAHCPVAVMRSKVDQPSADINMIVVMTADPDNEAVLEAAREAKLRQAPILAGR 187  
DB 129 LALSACCPVAVMRSKVDQPSADINMIVVMTADPDNEAVLEAAREAKLRQAPILAGR 188  
QY 188 PEELREIPD--GEFERRVQDMHRRHPDVRYPTTHTGTARFLADHDERVQLAVIGGE 244  
DB 189 QSGVGTGDDVRASLDRQPRYPDVYVQAAVHGEILDYLAGRGSVHVVLSASD 248  
QY 245 AGQLARLVGSPGHPFRHAECVLYV 270  
DB 249 QEHVEQLVGAPGNAVLQEAAGCTLLV 274

RESULT 3

AD137321 standard; protein; 268 AA.

AC AD137321;

DT 22-APR-2004 (first entry)

DE M. tuberculosis low oxygen induced antigen Rv3134c SEQ ID NO:42.

KM mycobacterial infection; vaccine; tuberculosis;

KM Mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;

XX low oxygen induced antigen.

XX Mycobacterium tuberculosis.

XX WO2004006952-A2.

XX 22-JAN-2004.

XX 08-JUL-2003; 2003WO-DK000477.

```
XX 13-JUL-2002; 2002DK-00001098.
PR (STAT-) STATENS SERUM INST.
PA
XX
PI Andersen P, Rosenkrands I, Stryhn A;
XX WPI; 2004-122778/12.
DR N-PSDB; ADI37366.
XX
PT Use of one or more polypeptides or their fragments, which are expressed
PT during the latent stage of the mycobacterial infection, and/or nucleic
PT acids encoding the polypeptides, for a therapeutic vaccine against
PT tuberculosis.
XX
PS Claim 3; SEQ ID NO 42; 76pp; English.
XX
CC The present invention describes polypeptides or their fragments, which
CC are expressed during the latent stage of a mycobacterial infection,
CC and/or nucleic acids encoding the polypeptides, which are useful for
CC creating a therapeutic vaccine against tuberculosis. Also described: (1)
CC a therapeutic vaccine against tuberculosis comprising one or more
CC polypeptides; (2) a method for treating an animal, including a human
CC being, with tuberculosis caused by virulent mycobacteria, e.g. by
CC Mycobacterium tuberculosis, M. africanum or M. bovis; (3) a method for
CC immunising an animal, including a human being, against tuberculosis
CC caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis
CC caused by virulent mycobacteria in an animal, including a human being;
CC (5) a method for diagnosing previous or ongoing infection with a virulent
CC mycobacterium; and (6) a method of diagnosing Mycobacterium tuberculosis
CC infection in a subject. The polypeptides have antibacterial activities,
CC and can be used in vaccines and in gene therapy. The polypeptides are
CC useful for the manufacture of a therapeutic vaccine for treating an
CC individual who is infected by a virulent mycobacterium, e.g. M.
CC tuberculosis, and who is not vaccinated with BCG against tuberculosis.
CC The present sequence represents a low oxygen induced antigen, which is
CC used in the exemplification of the present invention.
XX
SQ Sequence 268 AA;
Query Match 25.6%; Score 357.5; DB 8; Length 268;
Best Local Similarity 36.5%; Pred. No. 2.7e-28;
Matches 100; Conservative 35; Mismatches 110; Indels 29; Gaps 6;
QY 10 KTIIVGDSHAATTAALMGVDEAISRVAFLRVSVIKTPHSPDDYDRDLAHERSRLR 69
DB 8 RAVVVGIGSRPAATLAALMAVDEAVNRDIPRLRVVLDPSQLSAAEGCGGSARAAALHD 67
QY 70 AQSVAEAGKLVKLTETDIPRGAPFVLVEASRDAMICVSGVIGRVAASIIIGSTATELA 129
DB 68 ASRYVEATGQPVKIEITVLGCRPLTKMQSRSAAMLCVSGVGLD-HYRGRRGSVAATLA 126
QY 130 EKAHCPVAVMSKVDQPA--SDIMKIVVRMTDAPNEAVLEAAREALCLQAPILALGGR 187
DB 127 GSALCPVAVIHPSPAEPTTSQVSAVVAEV---DNGVLRHAPEAREALRGVPLRAVA-- 180
QY 188 PEULREIIPDG-----EPEERVDQMHHNRHPRVRYVPTTTTHGIALRPLADDERVQL 237
DB 181 -VHALETDDVDEQGRLAHLNLSRLAMTRLYPEVRDRALAGGSACRHLAANKRQQL 239
QY 238 AVIGGEGAGOLARLVGPSGHVFRHAECSVLVVR 271
DB 240 FVADSHSAHELCAVQPG-----CAVLTVR 264
RESULT 4
ADI37298
ID ADI37298 standard; protein; 295 AA.
XX
XX ADI37298;
XX
XX 22-APR-2004 (first entry)
```

```
DE M. tuberculosis low oxygen induced antigen Rv2050c SEQ ID NO:19.
XX mycobacterial infection; vaccine; tuberculosis;
XX Mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;
XX low oxygen induced antigen.
XX
OS Mycobacterium tuberculosis.
XX
PN WQ2004006952-A2.
XX
PD 22-JAN-2004.
XX
XX 08-JUL-2003; 2003WO-DK00477.
PR 13-JUL-2002; 2002DK-00001098.
XX (STAT-) STATENS SERUM INST.
XX
PI Andersen P, Rosenkrands I, Stryhn A;
XX WPI; 2004-122778/12.
DR N-PSDB; ADI37343.
XX
PT Use of one or more polypeptides or their fragments, which are expressed
PT during the latent stage of the mycobacterial infection, and/or nucleic
PT acids encoding the polypeptides, for a therapeutic vaccine against
PT tuberculosis.
XX
PS Claim 3; SEQ ID NO 19; 76pp; English.
XX
CC The present invention describes polypeptides or their fragments, which
CC are expressed during the latent stage of a mycobacterial infection,
CC and/or nucleic acids encoding the polypeptides, which are useful for
CC creating a therapeutic vaccine against tuberculosis. Also described: (1)
CC a therapeutic vaccine against tuberculosis comprising one or more
CC polypeptides; (2) a method for treating an animal, including a human
CC being, with tuberculosis caused by virulent mycobacteria, e.g. by
CC Mycobacterium tuberculosis, M. africanum or M. bovis; (3) a method for
CC immunising an animal, including a human being, against tuberculosis
CC caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis
CC caused by virulent mycobacteria in an animal, including a human being;
CC (5) a method for diagnosing previous or ongoing infection with a virulent
CC mycobacterium; and (6) a method of diagnosing Mycobacterium tuberculosis
CC infection in a subject. The polypeptides have antibacterial activities,
CC and can be used in vaccines and in gene therapy. The polypeptides are
CC useful for the manufacture of a therapeutic vaccine for treating an
CC individual who is infected by a virulent mycobacterium, e.g. M.
CC tuberculosis, and who is not vaccinated with BCG against tuberculosis.
CC The present sequence represents a low oxygen induced antigen, which is
CC used in the exemplification of the present invention.
XX
SQ Sequence 295 AA;
Query Match 20.5%; Score 286; DB 8; Length 295;
Best Local Similarity 29.6%; Pred. No. 8.7e-21;
Matches 85; Conservative 48; Mismatches 126; Indels 28; Gaps 8;
QY 12 IIVGIDGSHAATTAALMGVDEAISRVAFLRVSVIK-----PTHPSPDDY-----DRDL 60
DB 10 VVGVDSLSLSDAABACGATDAMRNIPLVVHVAVNDVATWPMPIPTETMGVQDEBGR 69
QY 61 AHAERSLREAQSAVEAAGKLVKLTETDIPRGAPFVLVEASRDAMICVSGVIGRVAASI 120
DB 70 QIVANNAVLAKEAAGADRKI-SVXSELVFTSPVPTWEISNEAEVMVLGSSGRLARGL 128
QY 121 LGSTATLEAKAHCPVAVMSKVDQPASDINW--IVVRMTDAPNEAVLEAAREALCLQ 178
DB 129 LGSVSSLVRRAGCPVAVIHSD--DAVLPDQHAPEVLGIGSPVSELATAVAFDEARRG 187
QY 179 APIALGGRPE-ELREIIPDGSEF-----RRVQDMHHRHPRVRYVPTTTTHGIALR 226
DB 188 VELTAVHMSDVEVLELPGIDFSAVQDEAELSLAERLAGHQERIPDPVSRVVVCDRPAR 247
```

OY	227	FLADHERVOLAVIGG-GEAGTAAITVSGSPVRHACSTLVYRR	272
Dd	248	KLVOKSASAQLVVVSHRGGLTGMLDSSNAVLHAARVPVIVARQ	294
RESULT	5		
XX	AY87878		
XX	AAy87878 standard; protein; 297 AA.		
XX	AAy87878;		
DT	06-OCT-2000	(first entry)	
XX	M. tuberculosis antigen TB32 protein.		
DE	Tuberculosis; TB; antigen; vaccine; diagnosis; somatic; tuberculostatic;		
RW	infection; interferon-gamma; IFN-gamma; protective immunity; therapy;		
KX	delayed type hypersensitivity response; TB32.		
XX	Mycobacterium tuberculosis.		
OS	MO200021983-A2.		
PV	20-APR-2000.		
XX	08-OCT-1999;	99WO-DK000538.	
PF	08-OCT-1998;	98DK-00001281.	
PR	21-JAN-1999;	99US-0116673P.	
XX	(START-) STATENS SERUM INST.		
PA	Andersen P, Weldingh K, Hansen CV, Florio W, Okkels LMM;		
PI	Skjot RLV, Rosenkrands I;		
DR	WPI: 2000-317931/27.		
N-	PSDB; AAA39572.		
PT	Novel polypeptide of somatic protein extract useful as vaccine against		
PT	viral Mycobacterium infection, isolated from cell wall, cell membrane		
XX	and cytoesol.		
XX	Claim 1; Page 106-107; 126pp; English.		
PS	This invention describes a novel polypeptide (PP) of somatic proteins		
XX	extract (I) which has tuberculostatic activity. (I) or their subsequence		
CC	has at least one of the following properties: (a) the PP induces an in		
CC	vitro recall response, or an in vitro response, during primary infection		
CC	with virulent Mycobacterium, determined by a release of interferon (IFN)-		
CC	gamma, (b) PP induces a protective immunity, determined by vaccinating an		
CC	animal with PP and an adjuvant, three times at two weeks intervals, (c)		
CC	PP induces an in vitro response, or in vitro recall response, determined		
CC	by release of IFN-gamma of at least 1000 pg/ml or 500 pg/ml,		
CC	respectively, from Peripheral Blood Mononuclear Cells (PBMC), withdrawn		
CC	from TB patients, or PPD positive individuals, 6 months after diagnosis,		
CC	(d) PP induces a specific antibody response in a TB patient, as		
CC	determined by enzyme linked immunosorbent assay (ELISA) technique or a		
CC	western blot. (e) PP induces a positive delayed type hypersensitivity		
CC	(DTH) response, determined by intradermal injection. (I) and (II) are		
CC	useful in preparing a prophylactic or therapeutic medicine as a vaccine		
CC	for induction of a protective or generation of an immune response in a		
CC	mammal against infection with a virulent Mycobacterium. (I) and (II) are		
CC	also useful as diagnostic reagent for the diagnosis of a virulent		
CC	Mycobacterium infection. The vaccine of the invention induces efficient		
CC	immunological memory, providing long term protection against TB. This		
CC	sequence represents a Microbacterium tuberculosis TB32 antigen described		
XX	in the invention		
SQ	Sequence 297 AA;		
Query Match	19.1%; Score 267; DB 3; Length 297;		
Best Local Similarity	29.3%; Pred. No. 8.4e-19;		
Matches	84; Conservative 42; Mismatches 131; Indels 30; Gaps 8;		

Oy	12	IIVGIDGSHAAITTAALMGDEAISAVALRLVSYIKP-----THPEP-----DDYDR	58
Oy	11	IIVGIDGSHAAITTAALMGDEAISAVALRLVSYIKP-----THPEP-----DDYDR	59
Dd	10	IIVGIDDSPPAAQVAVMAARADAEIKKILPLTLVHAVSPREVALTWLEVPRLPPEGLRWQODHGR	69
Oy	59	DLHAERSIPREKQSAVEAAGKLVKLETDIPPGAPFVLEASRDABEIMCVGSVIGRYAS	118
Dd	70	HL--IDDAKLVNVEQASLRAGP-PYVHSEIVPRAAVFTLVDSKALVLMVVGCLASGKWRPG	126
Oy	119	SILGSTATSIAEKAKCPVAMKSKVD-OPASDINMIVMTDARDNEAVLEAAREAKLR	177
Dd	127	RLTGSVSGGLRHAHCPRVYIHDEDSVMPHQQAPVLVGVGSSASASELATAIAFDEASRR	186
Oy	178	QAATLACGRPE-ELRELP-----DGEFERRVQDMHNHNRVRYPTITTHGTA	225
Dd	187	NVDVLVLAHMSVDVSEWPGIDWPATQSMAGQVLAERLAGQKQERYPVAAITRVVADQPA	246
Oy	226	RFLADDEREVQALVIGG-GEAGQLARLVGSPGHPFVRHAECSVLVVR	271
Dd	247	ROLVQSEERQVLVVQSGRGSGYAGMLVGSGETVAQALARTPVIVAR	253
RESULT 6			
ABG30907			
ID	ABG30907	standard; protein; 297 AA.	
XX	ABG30907;		
AC			
XX	21-OCT-2002 (first entry)		
DT			
XX			
DE	Mycobacterium tuberculosis dormancy-induced protein Rv2623.		
XX			
KM	Anti-mycobacterial; Mycobacterium tuberculosis; strain H37Rv;		
KM	non-oxygen limiting; hypoxic stationary; hypoxic growth phase;		
KM	tuberculosis; dormant mycobacterial infection; dormancy-induced; Rv2623;		
XX	tuberculostatic; vaccine.		
XX			
OS	Mycobacterium tuberculosis.		
XX			
PN	WO200248391-A2.		
PD	20-JUN-2002.		
XX			
XX	11-DEC-2001; 2001WO-EP014551.		
PF			
XX	13-DEC-2000; 2000GB-00030368.		
PR			
XX	(MOLB-) INST MOLECULAR & CELL BIOLOGY.		
XX			
PA	Dick T, Calvin BKK;		
PI			
XX	WPI; 2002-583530/62.		
DR	N-PSDB; ABR93546.		
XX			
PT	Identifying an anti-mycobacterial agent that modulates		
PT	activity/expression of a protein expressed by Mycobacterium, involves		
PT	monitoring the effect of an agent on the activity/expression of the		
PT	protein or polynucleotide/vector encoding it.		
XX			
PS	Example 2; Page 44-45; 56pp; English.		
XX			
CC	The present invention relates to a new method for the identification of		
CC	anti-mycobacterial agent that modulates activity and/or expression of		
CC	protein expressed by Mycobacterium in non-oxygen limiting or hypoxic		
CC	stationary, hypoxic growth phase. The method involves contacting a test		
CC	agent and protein, and monitoring the effect of agent on activity/		
CC	expression of the protein. The methods of the invention are useful for		
CC	treating a human or animal body by therapy, in a diagnostic method		
CC	practised on the human or animal body, and for manufacturing medicament		
CC	for diagnosis, prophylaxis or treatment of mycobacterial infection,		
CC	especially tuberculosis. The invention is also useful for in vitro or in		
CC	vivo diagnosing of mycobacterial infection in a sample. The invention can		
CC	be used for diagnosing a dormant mycobacterial infection. The present		

CC amino acid sequence represents the *Mycobacterium tuberculosis* strain  
 CC H37Rv dormancy-induced RV2623 protein of the invention  
 XX  
 SO Sequence 297 AA;

Query Match 19.1%; Score 267; DB 5; Length 297;  
 Best Local Similarity 29.3%; Pred. No. 8.4e-19;  
 Matches 84; Conservative 42; Mismatches 131; Indels 30; Gaps 8;

QY IIVGIDGSHAATTAALMGVDEAISRAPPLRLVSVIKP-----THSPD-----DDYDR 58  
 DB IIVGIDGSPAAQAVRAVMAARDSELRKIPLTIVHAVSPVAVTWLEVPGLRWQDQHR 69  
 QY DLAAHSELRSAQSAVEAAGKLVKIETDIPRGAPGVLVEASRDAMI CVSGVIGRYAS 118  
 DB HL--IDDLAKVVEQASLRAGP-PTVHSEIVPAAAVPTLVDSKDAVLVVGCLSGGRWPG 126  
 QY SIIGSTATELAEKACPVAVWRSKYD-QPASDINIVVVRMTDAPNEAVLEAAREACTR 177  
 DB RLTSVSSSGLLRHACPVVLIHDEDSVMPHPQAPVLGVGSSASSELATATAFDEASRR 186  
 QY QAPILALGRRPE-ELREIP-----DGEFERRVQDMHRRHDPVVRVPTTHTGIA 225  
 DB NVDLVLAHMSQVDVSEWPGCIDMPATQSMAEQVLAERLAGHQBRYENVAITRVVRDQPA 246  
 QY RFLADHDERVOLAVIGG-GEAGQLARLVPSGHPVFRHAEGSVLVVR 271  
 DB RQLVGRSEBAQLVNVGSRGRGVAGMLVGSVGETVAQALARPVIYAR 293

RESULT 7  
 AAB79640  
 ID AAB79640 standard; protein; 301 AA.

AC AAB79640;  
 XX  
 DT 30-APR-2001 (first entry)  
 XX

DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:14.

XX  
 KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
 KW fine chemical production; microorganism; organic acid; nucleoside;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;  
 KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
 KW carbohydrate; aromatic compound; cofactor; polypeptide; enzyme.

OS Corynebacterium glutamicum.

XX WO200100843-A2.

XX  
 PD 04-JAN-2001.

XX  
 PP 23-JUN-2000; 2000WO-IB000923.

XX 25-JUN-1999; 99US-0141031P.  
 PR 01-JUL-1999; 99US-01030476.  
 PR 02-JUL-1999; 99US-0142101P.  
 PR 08-JUL-1999; 99DE-01031415.  
 PR 08-JUL-1999; 99DE-01031418.  
 PR 08-JUL-1999; 99DE-01031419.  
 PR 08-JUL-1999; 99DE-01031420.  
 PR 08-JUL-1999; 99DE-01031424.  
 PR 08-JUL-1999; 99DE-01031428.  
 PR 08-JUL-1999; 99DE-01031434.  
 PR 08-JUL-1999; 99DE-01031435.  
 PR 08-JUL-1999; 99DE-01031443.  
 PR 08-JUL-1999; 99DE-01031453.  
 PR 08-JUL-1999; 99DE-01031457.  
 PR 08-JUL-1999; 99DE-01031465.  
 PR 08-JUL-1999; 99DE-01031478.  
 PR 08-JUL-1999; 99DE-01031510.  
 PR 08-JUL-1999; 99DE-01031541.  
 PR 08-JUL-1999; 99DE-01031573.

PR 08-JUL-1999; 99DE-01031592.  
 PR 08-JUL-1999; 99DE-01031632.  
 PR 08-JUL-1999; 99DE-01031634.  
 PR 08-JUL-1999; 99DE-01031636.  
 PR 09-JUL-1999; 99DE-01032125.  
 PR 09-JUL-1999; 99DE-01032126.  
 PR 09-JUL-1999; 99DE-01032130.  
 PR 09-JUL-1999; 99DE-01032186.  
 PR 09-JUL-1999; 99DE-01032206.  
 PR 09-JUL-1999; 99DE-01032227.  
 PR 09-JUL-1999; 99DE-01032228.  
 PR 09-JUL-1999; 99DE-01032229.  
 PR 09-JUL-1999; 99DE-01032230.  
 PR 14-JUL-1999; 99DE-01032922.  
 PR 14-JUL-1999; 99DE-01032926.  
 PR 14-JUL-1999; 99DE-01032928.  
 PR 14-JUL-1999; 99DE-01033004.  
 PR 14-JUL-1999; 99DE-01033005.  
 PR 14-JUL-1999; 99DE-01033006.  
 PR 12-AUG-1999; 99US-0148613P.  
 PR 27-AUG-1999; 99DE-01040764.  
 PR 27-AUG-1999; 99DE-01040765.  
 PR 27-AUG-1999; 99DE-01040766.  
 PR 27-AUG-1999; 99DE-01040832.  
 PR 31-AUG-1999; 99DE-01041378.  
 PR 31-AUG-1999; 99DE-01041379.  
 PR 31-AUG-1999; 99DE-01041380.  
 PR 31-AUG-1999; 99DE-01041394.  
 PR 31-AUG-1999; 99DE-01041396.  
 PR 03-SEP-1999; 99DE-01042076.  
 PR 03-SEP-1999; 99DE-01042077.  
 PR 03-SEP-1999; 99DE-01042079.  
 PR 03-SEP-1999; 99DE-01042086.  
 PR 03-SEP-1999; 99DE-01042087.  
 PR 03-SEP-1999; 99DE-01042088.  
 PR 03-SEP-1999; 99DE-01042095.  
 PR 03-SEP-1999; 99DE-01042124.  
 PR 03-SEP-1999; 99DE-01042129.  
 PR 09-MAR-2000; 2000US-0187970P.  
 XX  
 BA (BADI ) BASF AG.  
 XX  
 PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;  
 XX  
 PI WPI; 2001-137957/14.  
 DR N-PSDB; AAF71759.  
 XX  
 XX Nucleic acids from *Corynebacterium glutamicum* encoding metabolic pathway  
 PT proteins, useful for producing fine chemicals in microorganisms,  
 PT including organic acids, nonproteinogenic amino acids, and purine and  
 PT pyrimidine bases.  
 XX  
 PS Claim 20; Page 174-175; 1737pp; English.

XX AAF71753 to AAF72330 encode the *Corynebacterium glutamicum* metabolic  
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum MP  
 CC nucleic acids are useful for the production of fine chemicals in  
 CC microorganisms, including organic acids, nonproteinogenic amino acids,  
 CC purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated  
 CC and unsaturated fatty acids, diols, carbohydrates, aromatic compounds,  
 CC vitamins, cofactors, polypeptides and enzymes

XX Sequence 301 AA;

Query Match 19.0%; Score 266; DB 4; Length 301;  
 Best Local Similarity 28.8%; Pred. No. 1.1e-18;  
 Matches 85; Conservative 45; Mismatches 131; Indels 34; Gaps 8;

QY 8 TMKTIIVGIDGSHAATTAALMGVDEAISRAPPLRLVSVIKP-----PDDYDR 58  
 DB 3 TEDIVVAVDSDSKQAVRAVMAANTANKRGIPRLIAS--SYTMQFLYAGMVPVPGQLFD 60  
 QY 59 DL-AHAERSLREASAVEAAGKLVKIETDIPRGAPGVLVEASRDAMI CVSGVIGRYA 117

Db 61 DLQEALEKINEADIAHEVAPEIKIGHTIAGEPIDMLEMSPDATIIVWGSSGLGELS 120  
 Qy 118 SSIGSTATLEAKHCFAVAMR--SKVDOPASDINMIIVMTDAPDNEAVLEAAREAK 175  
 Db 121 GMMWGSVGAIVVSHAKCPVVVVRSDSAVNEP-SKGPVVVGVDSSEVSOQATEYAFAEAE 179  
 Qy 176 LR-----QAPILALGGRPEBELREIPDGEFF--RRVODMHHRRPDVRYVP 217  
 Db 180 ARGAEIVAVHTWMDMOVASLAGLAAAOQOWDEVERQOTDMLIERLAPLVEKYPSTVVK 239  
 Qy 218 ITHHTGIARFLADDERVQLAVIGG-GEAGQLARLVGSGHPRVRRHACSVLYVR 271  
 Db 240 IITRDPRVRLALAEASENAOQLLVGSHGRGFKGMILGSTSRALLQSAFCPMVYR 294

RESULT 8  
 AAU71872  
 ID AAU71872 standard; protein; 301 AA.  
 XX AAU71872;  
 AC  
 XX 26-FEB-2002 (first entry)  
 DT  
 XX C. glutamicum metabolic pathway protein encoded by gene #7.  
 DE  
 XX Metabolic pathway protein; MP; lysine biosynthesis pathway;  
 KW methionine biosynthesis pathway; large-scale production of fine chemical;  
 KM Corynebacterium diphtheriae; diphtheria.  
 XX  
 OS Corynebacterium glutamicum.  
 KM  
 XX MO200166573-A2.  
 PN  
 XX 13-SEP-2001.  
 PD  
 XX 22-DEC-2000; 2000MO-IB002035.  
 PF  
 XX 09-MAR-2000; 2000US-0187970P.  
 PR 23-JUN-2000; 2000US-00606740.  
 XX  
 XX (BADI ) BASF AG.  
 PA  
 XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
 PI Kim J, Lee H, Hwang B;  
 XX  
 XX WPI; 2001-582269/65.  
 DR N-PSDB; AAS96082.  
 DR  
 PT Nucleic acids encoding metabolic pathway proteins from Corynebacterium  
 PT glutamicum, useful for producing methionine and lysine in Corynebacterium  
 PT and Brevibacterium.  
 PT  
 XX  
 PS Disclosure; Page 169-170; 316pp; English.  
 XX  
 XX The present invention relates to the isolation of novel Corynebacterium  
 CC glutamicum genes (AAS96073-AAS96132) encoding metabolic pathway (MP)  
 CC proteins. The metabolic pathway proteins of the invention include enzymes  
 CC involved in the lysine and methionine biosynthetic pathways. The  
 CC polynucleotide sequences of the invention can be used for the large-scale  
 CC production and/or modulation of expression of fine chemicals such as  
 CC lysine and methionine. The sequences of the invention may be used to  
 CC identify C. glutamicum and related organisms e.g. C. diphtheriae in a  
 CC subject to detect diphtheria. AAU71863-AAU71922 represent the novel C.  
 CC glutamicum metabolic pathway proteins of the invention  
 CC  
 XX  
 SQ Sequence 301 AA;

Query Match 19.0%; Score 266; DB 4; Length 301;  
 Best Local Similarity 28.8%; Pred. No. 1,1e-18;  
 Matches 85; Conservative 45; Mismatches 111; Indels 34; Gaps 8;

8 TMTIIVGIDGSHAATRALMGVDEAISRAVPLRLVSVIKPTPS-----PDDYDR 58

Db 3 TEDIVVAVVDGSPASAKOAVMANTANKRGIPRLALS--SYTFPOFLYAGMVPPELFD 60  
 Qy 59 DL-AHAERSRREAOASAVEAGKLVKLETDP-PRRPAGVLEVASRDAMEICVSGVIGRYA 117  
 Db 61 DLQEALEKINEADIAHEVAPEIKIGHTIAGEPIDMLEMSPDATIIVWGSSGLGELS 120  
 Qy 118 SSIGSTATLEAKHCFAVAMR--SKVDOPASDINMIIVMTDAPDNEAVLEAAREAK 175  
 Db 121 GMMWGSVGAIVVSHAKCPVVVVRSDSAVNEP-SKGPVVVGVDSSEVSOQATEYAFAEAE 179  
 Qy 176 LR-----QAPILALGGRPEBELREIPDGEFF--RRVODMHHRRPDVRYVP 217  
 Db 180 ARGAEIVAVHTWMDMOVASLAGLAAAOQOWDEVERQOTDMLIERLAPLVEKYPSTVVK 239  
 Qy 218 ITHHTGIARFLADDERVQLAVIGG-GEAGQLARLVGSGHPRVRRHACSVLYVR 271  
 Db 240 IITRDPRVRLALAEASENAOQLLVGSHGRGFKGMILGSTSRALLQSAFCPMVYR 294

RESULT 9  
 AAU71865  
 ID AAU71865 standard; protein; 301 AA.  
 XX AAU71865;  
 AC  
 XX 26-FEB-2002 (first entry)  
 DT  
 XX C. glutamicum metabolic pathway protein encoded by RXA00657 gene.  
 DE  
 XX Metabolic pathway protein; MP; lysine biosynthesis pathway;  
 KW methionine biosynthesis pathway; large-scale production of fine chemical;  
 KM Corynebacterium diphtheriae; diphtheria.  
 XX  
 OS Corynebacterium glutamicum.  
 KM  
 XX MO200166573-A2.  
 PN  
 XX 13-SEP-2001.  
 PD  
 XX 22-DEC-2000; 2000MO-IB002035.  
 PF  
 XX 09-MAR-2000; 2000US-0187970P.  
 PR 23-JUN-2000; 2000US-00606740.  
 XX  
 XX (BADI ) BASF AG.  
 PA  
 XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
 PI Kim J, Lee H, Hwang B;  
 XX  
 XX WPI; 2001-582269/65.  
 DR N-PSDB; AAS96075.  
 DR  
 PT Nucleic acids encoding metabolic pathway proteins from Corynebacterium  
 PT glutamicum, useful for producing methionine and lysine in Corynebacterium  
 PT and Brevibacterium.  
 PT  
 XX  
 PS Claim 24; Page 150-151; 316pp; English.  
 XX  
 XX The present invention relates to the isolation of novel Corynebacterium  
 CC glutamicum genes (AAS96073-AAS96132) encoding metabolic pathway (MP)  
 CC proteins. The metabolic pathway proteins of the invention include enzymes  
 CC involved in the lysine and methionine biosynthetic pathways. The  
 CC polynucleotide sequences of the invention can be used for the large-scale  
 CC production and/or modulation of expression of fine chemicals such as  
 CC lysine and methionine. The sequences of the invention may be used to  
 CC identify C. glutamicum and related organisms e.g. C. diphtheriae in a  
 CC subject to detect diphtheria. AAU71863-AAU71922 represent the novel C.  
 CC glutamicum metabolic pathway proteins of the invention  
 CC  
 XX  
 SQ Sequence 301 AA;

Query Match 19.0%; Score 266; DB 4; Length 301;

Best Local Similarity 28.8%; Pred. No. 1.1e-18;  
Matches 85; Conservative 45; Mismatches 131; Indels 34; Gaps 8;

```

Qy      8  FMKTTIIVGDSHAITLALMGVDEIASVAPRLTYSVILKPTHS-----PDDYR  58
Dy      3  TEDIIVAVDGDSDASQOAVRMAANTANKGIFRLMS--SYMPQFLYAEGVPPQSLFD  60
Qy      59  DL-AHAERSLREAQSAVEAAGKLVKIETDI PRGPAPVLVEASRDAMI CVSGVIGRYA  117
Dy      61  DLQAELEKINEARDIAHEVAPEIKGHTIAEGSPIDMLLEMS PDATMI VMGSRGIGLS  120
Qy      118  SSIIIGSTATELAEKXHCPCVAVNR--SKYDQPSADIMWIYVRMTDAEDNEAVLEYARBEAK  175
Dy      121  GMVGSVSGAVVSHAKCPVVVVRBSAVNEI--SKYCPVVAVGVDSSEVSQDTEYAAREAE  179
Qy      176  LR-----CAPILALGGRPEELREIPDGEFE--RAYVDMMHHRPDRVVR  217
Dy      180  ARGAEIVAVHTIMDMQVQASLGLAAAOQOMPEVEHQQIDMLIERLAPLVEKIPSTVVK  239
Qy      218  ITTHTGIAFLADHDREVVLAVICG-GEAGQILARLVGSGHPVFRIAEGSVLVR  271
Dy      240  IITRDPRVALAEASNAQLLVVSGHSGRGFKMLIGTSRLLQAGPCPMVVR  294

```

RESULT 10  
AAG92991  
ID AAG92991 standard; protein; 301 AA.

PT Novel polynucleotides derived from *Corynebacterium* bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analyzing  
PT expression profile or pattern of a gene and identifying homologous gene.  
XX  
PS Claim 17; SEQ ID NO 6745; 246bp +Sequence Listing; English.

CC The present invention provides a number of nucleotide and protein  
CC sequences from the *Corynebacterium* *Corynebacterium glutamicum*. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of *Corynebacterium* *Corynebacterium*, measuring expression amount and analysing  
CC the expression profile or expression pattern of a gene derived from  
CC *Corynebacterium* *Corynebacterium*, and identifying a homologue of a gene derived from  
CC *Corynebacterium* *Corynebacterium*. *Corynebacterium* bacteria are useful for producing amino  
CC acids, nucleic acids, vitamins, saccharides and organic acids.  
CC particularly L-lysine. The present sequence is a protein described in the  
CC exemplification of the invention. Note: The sequence data for this patenting

CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from the European Patent Office  
XX  
50 Sequence 301 AA;

Query Match 19.0%; Score 266; DB 4; Length 301;

Matches	85;	Conservative	45;	Mismatches	131;	Indels	34;	Gaps	8;
---------	-----	--------------	-----	------------	------	--------	-----	------	----

```

QY      8  TMTITIVGIDGSHAITALMGVDAISRPAIVSYIKRTHS-----PDDYDR  58
           :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      3  TEDIIVAADSDASQAVRMAMANTANKGILPLRLAS--STTTFQFLYAGSMPPDELF  60
           :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY      59  DL-AHAERSLRKXASAVEAAGCLVYIETDIPRGPAPVLYEASHDAEMI CVGSVIGRYA  112
           :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      61  DLOVALTEKINEARDIAHVAPEIKIGHTIAEGSPIDWLEMSPDATMI VMGSGIGLGS  120
           :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY      118  SSLGSTATLEAKAKCPVAVNR--SKYDQASDINMTVVRMTAPRNEAVLEAAPEAK  175
           :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      121  GMMGSGVSAQVASHAKCPVVVREDASANE- SKYGPVVGVDSEVSQATEYAFPEAE  179
           :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY      176  LR-----CAPILALGGRPEELREIPDEGFE---RRVDMHNRHDPVAVVP  212
           :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      180  ARGAEILVAHTMTMDQVQASLAGLAQAQOMDEVERQOTDYLIRLRALIVKEKPSVTVKK  239
           :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY      218  ITTHTGIAPLADHDREVOLAVTGG-GEAQGLARLVPSGHPVRNAECSVLVVR  271
           :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      240  IITRDRPVALLAEASNAQLLVGVGHGGGFGFKMILGSTSPALQASAPCPMMVVR  294
           :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 11  
ABG80324  
ID ABG80324 standard; protein; 301 AA

KM Metabolic pathway protein; MP; cell metabolism; amino acid; vitamin;  
KM cofactor; nucleoside; nucleoside; trehalose; fine chemical production  
KM organic acid; non-proteinogenic amino acid; purine base; carboxylate  
KM pyrimidine base; lipid; unsaturated fatty acid; diol; polyketide;  
KM aromatic compound; food industry; animal feed; cosmetic industry;  
KM pharmaceutical industry; enzyme.

PT New metabolic pathway genes of *Corynebacterium glutamicum* for producing  
PT fine chemicals, e.g. lipids, (un)saturated fatty acids, vitamins,  
PT cofactors or enzymes used in food, feed, cosmetics or pharmaceutical  
PT industries.

CC The present invention relates to the isolation of *Corynebacterium*  
CC glutamicum metabolic pathway (MP) proteins, and the polynucleotide



CC sequences encoding them. The MP proteins are enzymes involved in the  
CC metabolism of molecules important for the normal functioning of cells  
CC (e.g. amino acids, vitamins, cofactors, nucleosides and nucleosides, or  
CC trehalose). The polynucleotide sequences encoding the MP proteins are  
CC useful for producing fine chemicals, particularly organic acids, non-  
CC proteinogenic amino acids, purine and pyrimidine bases, nucleosides,  
CC nucleotides, lipids, (un)saturated fatty acids, diols, carbohydrates,  
CC aromatic compounds, vitamins, cofactors, polyketides and enzymes. The  
CC fine chemicals are useful in the food, animal feed, cosmetic or  
CC pharmaceutical industries. ABG60321-ABG80343 represent the C. glutamicum  
CC MP proteins of the invention

XX Sequence 301 AA;

SO Query Match 19.0%; Score 266; DB 5; Length 301;

Best Local Similarity 28.8%; Pred. No. 1.1e-18;

Matches 85; Conservative 45; Mismatches 131; Indels 34; Gaps 8;

QY 8 TMKTIIVGIDGSHAAITTAALMGVDEAISRANPRLRVSVIKPTNPS-----PDDYDR 58  
DB 3 TEDIVVAVDGSDASKQAVRMAANTANKRGIPRLAS--SYTMOFLYAEQWVPOELFD 60  
QY 59 DL-AHAERSLREASQSAVEAGKLVKIEETDIPRGAPVLYEASRDMEICVSGVIGRYA 117  
DB 61 DLQAELEKINEKRDIAHEVAPEIKIGHTIAEGSPIDMLENSPDATMYVMSRGIGLS 120  
QY 118 SSIIGSTATLEAKNACPAVAMR--SKVDQPSADIMVIWMTDAPDNEAVLEYAAREAK 175  
DB 121 GWMVGSVGAIVSHAKCPVVVVRSDAVNED--SKYGPVVGVDSSEVQATEYAFPAEAE 179  
QY 176 LR-----QAPILALGGRPEELRIIPDGEFE--RRVQDMHNRHPRVRYR 217  
DB 180 ARGAEIVAVHTMDMQVQASLAGLAAQQQWDEVERQOTDMLERLAPLVEKYPSTVYVK 239  
QY 218 IITHTGIARPLADDERVOLAVIGG-GEAGQLARLVGPSGHPVFRHAECSVLVVR 271  
DB 240 IITDRPRVALAEASNAQLLVGSHGRGFKMGLGSTSRALLQSAFCPPMAYVR 294

RESULT 12

ID AEB15399 standard; protein; 301 AA.

XX AEB15399;

AC 22-SEP-2005 (first entry)

XX C glutamicum metabolic pathway regulatory (MR) protein SeqID314.

XX metabolism, microorganism detection, microorganism identification;

XX genome; evolution; protein production.

XX Corynebacterium glutamicum.

OS US2005153402-A1.

XX 14-JUL-2005.

PD 06-DEC-2004; 2004US-0006098.

XX 25-JUN-1999; 99US-0141031P.

XX 01-JUL-1999; 99DE-01030476.

XX 01-JUL-1999; 99US-0142690P.

XX 08-JUL-1999; 99DE-01031419.

XX 08-JUL-1999; 99DE-01031420.

XX 09-JUL-1999; 99DE-01032122.

XX 09-JUL-1999; 99DE-01032128.

XX 09-JUL-1999; 99DE-01032134.

XX 09-JUL-1999; 99DE-01032207.

XX 14-JUL-1999; 99DE-01033003.

XX 27-AUG-1999; 99US-0151251P.

XX 31-AUG-1999; 99DE-01041390.

PR 03-SEP-1999; 99DE-01042088.  
PR 03-SEP-1999; 99DE-01042124.  
PR 23-JUN-2000; 2000US-00602874.

XX (BADI ) BASF AG.

PI Pompejus M, Kroger B, Schroder H, Zelder O, Haberhauer G;

XX WPI; 2005-496831/50.

DR N-PSDB; AEB15398.

XX An isolated metabolic pathway regulatory polypeptide from Corynebacterium

PT glutamicum, its portion, or its variant, useful for diagnosing presence

PS Claim 20; SEQ ID NO 314; 65pp; English.

CC This invention relates to novel isolated metabolic pathway regulatory  
CC polypeptides from Corynebacterium glutamicum, designated MR proteins, and  
CC the DNA sequences which encode them. The invention is useful in  
CC identification of C glutamicum and related organisms, mapping of genomes  
CC of organisms related to C glutamicum, identification and localization of  
CC C glutamicum sequences of interest, evolutionary studies, determination  
CC of MR protein regions required for function, modulation of MR protein  
CC activity and modulation of cellular production of desired compounds such  
CC as a fine chemical. The present sequence is that of a novel isolated  
CC metabolic pathway regulatory polypeptide of the invention from  
CC Corynebacterium glutamicum. Note: This sequence does not appear in the  
CC specification but was obtained in electronic format directly from USPTO.

XX Sequence 301 AA;

SO Query Match 19.0%; Score 266; DB 9; Length 301;

Best Local Similarity 28.8%; Pred. No. 1.1e-18;

Matches 85; Conservative 45; Mismatches 131; Indels 34; Gaps 8;

QY 8 TMKTIIVGIDGSHAAITTAALMGVDEAISRANPRLRVSVIKPTNPS-----PDDYDR 58  
DB 3 TEDIVVAVDGSDASKQAVRMAANTANKRGIPRLAS--SYTMOFLYAEQWVPOELFD 60  
QY 59 DL-AHAERSLREASQSAVEAGKLVKIEETDIPRGAPVLYEASRDMEICVSGVIGRYA 117  
DB 61 DLQAELEKINEKRDIAHEVAPEIKIGHTIAEGSPIDMLENSPDATMYVMSRGIGLS 120  
QY 118 SSIIGSTATLEAKNACPAVAMR--SKVDQPSADIMVIWMTDAPDNEAVLEYAAREAK 175  
DB 121 GWMVGSVGAIVSHAKCPVVVVRSDAVNED--SKYGPVVGVDSSEVQATEYAFPAEAE 179  
QY 176 LR-----QAPILALGGRPEELRIIPDGEFE--RRVQDMHNRHPRVRYR 217  
DB 180 ARGAEIVAVHTMDMQVQASLAGLAAQQQWDEVERQOTDMLERLAPLVEKYPSTVYVK 239  
QY 218 IITHTGIARPLADDERVOLAVIGG-GEAGQLARLVGPSGHPVFRHAECSVLVVR 271  
DB 240 IITDRPRVALAEASNAQLLVGSHGRGFKMGLGSTSRALLQSAFCPPMAYVR 294

RESULT 13

ID ABU25937 standard; protein; 300 AA.

XX ABU25937;

XX 19-JUN-2003 (first entry)

XX Protein encoded by prokaryotic essential gene #11464.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Corynebacterium diptheriae.

XX WO200277183-A2.

PD 03-OCT-2002.  
 XX 21-MAR-2002; 2002WO-US009107.  
 PF 21-MAR-2001; 2001US-00815242.  
 XX 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Hasebeck R, Ohlsen KL, Zykkind JW,  
 P1 Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,  
 P1 WPI; 2003-029926/02.  
 DR N-PSDB; ACN29807.  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 53861; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 XX  
 SO Sequence 300 AA;  
 Query Match 17.8%; Score 249; DB 6; Length 300;  
 Best Local Similarity 28.5%; Pred. No. 6.4e-17;  
 Matches 84; Conservative 46; Mismatches 123; Indels 42; Gaps 10;  
 Oy 12 IIVGIDGSHAATLAWGVDEAISRAVPLRLVSVIKPHPPDD-YDRDLHAERSLAEBA 70  
 Db 7 VVVAVDGSEASONARMAANTANKGCVPLRLA---SYTFMDFIYAGMPPQGLFEBL 62  
 Oy 71 QS-----AVEAGKL-----VKIETDIPRGPAGVLTVEASRDAMICVSGVIGRYASS 119  
 Db 63 QSEMTDMIEARVVAHEVAPDKIGVYVAESGPIMLDMSDDVMIMVMSGSGIGLGS 122  
 Oy 120 ILGSTATLAEKACFVAVMRSKVDQPASDINW---IYVRMTADPNEAVLEAYAREKL 176  
 Db 123 VMGSVAIVSHADCPVVVVR--DNHYTEIKYGPVVVVGSDVDSORATEFAFEAAQA 180

Oy 177 R-----CAPITALGRPEELREIPDGE----FERRVODWHRRHDPVRYVP 217  
 Db 181 RGAKLVAIHTWMDMQVQASLAGLAAQOEW-ETIEKEQTLLKDRLOPLERFPDVEVM 239  
 Oy 218 ITHHTGAPRLADDERVOQLAVTGG-GEAGQLARLVNPSGSHPRVRHAECSGLVVR 271  
 Db 240 VITDRPVRRLIEDCAHNAQLLVGSHGRGFRGMLLASTSRALLQSAPCPMVVR 294  
 RESULT 14  
 AAU62080  
 ID AAU62080 standard; protein; 316 AA.  
 XX  
 AC AAU62080;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #22976.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN MO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PP 20-APR-2001; 2001WO-US012865.  
 XX  
 PR 21-APR-2000; 2000US-0199047P.  
 PR 02-JUN-2000; 2000US-0208841P.  
 PR 07-JUL-2000; 2000US-0216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skelky YM, Persing DH, Mitcham JL, Wang SS, Bhactia A;  
 P1 L'maisemeuve J, Zhang Y, Jen S, Carter D;  
 DR N-PSDB; AAS59623.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris.  
 XX  
 PS Example 1; SEQ ID NO 23275; 1069pp; English.  
 XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC *P. acnes* is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of *P. acnes* in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for *P. acnes* proteins. These antibodies can be used to  
 CC downregulate expression and activity of *P. acnes* polypeptides and  
 CC therefore treat *P. acnes* infections. The antibodies may also be used as  
 CC diagnostic agents for determining *P. acnes* presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 XX  
 SO Sequence 316 AA;



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 05:11:18 ; Search time 10.8216 Seconds  
(without alignments)  
2418.401 Million cell updates/sec

Title: US-10-617-038-27  
Perfect score: 1397  
Sequence: 1 MSGSGEPTMKTIIVGIDGSH.....GPSGHPVFRHAECSVLVRR 272

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1397	100.0	272	2 G70572	hypothetical prote
2	416	29.8	279	2 G70942	hypothetical prote
3	357.5	25.6	268	2 G70645	hypothetical prote
4	286.5	21.4	294	2 A70942	hypothetical prote
5	286	20.5	295	2 C70759	hypothetical prote
6	267	19.1	297	2 F70572	hypothetical prote
7	258.5	18.5	317	2 B70758	hypothetical prote
8	243.5	17.4	288	2 T37029	hypothetical prote
9	230	16.5	283	2 T36962	hypothetical prote
10	224.5	16.1	301	2 T37031	hypothetical prote
11	220	15.7	294	2 T36949	conserved hypotet
12	186	13.3	328	2 T36963	hypothetical prote
13	160.5	11.5	201	2 T37036	hypothetical prote
14	149	10.7	152	2 T36954	hypothetical prote
15	122.5	8.8	174	2 C86231	hypothetical prote
16	119	8.5	170	2 C64366	hypothetical prote
17	116.5	8.3	162	2 F64918	hypothetical prote
18	115.5	8.3	150	2 G84340	hypothetical prote
19	114	8.2	150	2 A75310	hypothetical prote
20	112.5	8.1	139	2 A80454	conserved hypotet
21	112	8.0	270	2 G69469	conserved hypotet
22	109.5	7.8	166	2 B89955	conserved hypotet
23	109	7.8	941	2 A70722	probable gcwb prot
24	107.5	7.7	1997	2 T30874	virginiamycin S sy
25	107	7.7	143	2 AC1497	conserved hypotet
26	106.5	7.6	143	2 B80349	conserved hypotet
27	105.5	7.6	462	2 T28699	hypothetical prote
28	104.5	7.5	147	2 H67082	conserved hypotet
29	103	7.4	143	2 AD1139	conserved hypotet

30	102.5	7.3	271	2 T35065	tryptophan synthas
31	102	7.3	591	2 F64334	acetolactate synth
32	101	7.2	141	2 AH2004	hypothetical prote
33	101	7.2	171	2 H98097	conserved hypotet
34	99.5	7.1	146	2 B70560	hypothetical prote
35	99.5	7.1	155	2 C86634	conserved hypotet
36	99.5	7.1	422	2 G70707	probable puri prot
37	99	7.1	154	2 AD1272	hypothetical prote
38	98.5	7.1	184	2 C69829	conserved hypotet
39	98.5	7.1	406	2 C70640	hypothetical prote
40	98.5	7.1	421	2 E83459	conserved hypotet
41	98	7.0	145	2 A86854	conserved hypotet
42	97	6.9	150	2 F95233	universal stress p
43	97	6.9	152	2 AF1634	hypothetical prote
44	96.5	6.9	190	2 T30796	double-stranded RN
45	96.5	6.9	190	2 G42508	E31 protein - vacc

ALIGNMENTS

```
RESULT 1
G70572      hypothetical protein Rv2624c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C:Accession: G70572
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churches, G.; Harris, D.; Gordon, S.;
Rajandream, M.A.; Rogers, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70572
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Residues: 1-272 <COL>
A:Residues type: DNA
A:Cross-references: UNIPROT:O06188; UNIPARC:UPI00000D5EAL; GB:Z95387; GB:AL123456; NID:G
A:Experimental source: strain H37Rv
C:Genetic8:
A:Gene: Rv2624c

Query Match      100.0%; Score 1397; DB 2; Length 272;
Best Local Similarity 100.0%; Pred. No. 3.8e-99;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 MSGSGEPTMKTIIVGIDGSHAITAALMGVDEAISRAVPLRLSVIRTPHSPDDYRDL 60
DB      1 MSGSGEPTMKTIIVGIDGSHAITAALMGVDEAISRAVPLRLSVIRTPHSPDDYRDL 60

OY      61 AAHARSIREAQSAAVEAAKVKIETDIPRGAGVLYEASRDAMICVSGIGRYASSI 120
DB      61 AAHARSIREAQSAAVEAAKVKIETDIPRGAGVLYEASRDAMICVSGIGRYASSI 120

OY      121 LGSTATLAEAKHCPVAVMSKVDQPSADINMIIVRMTPDPADEAVLEVAAREAKLRQAP 180
DB      121 LGSTATLAEAKHCPVAVMSKVDQPSADINMIIVRMTPDPADEAVLEVAAREAKLRQAP 180

OY      181 ILALGSRPEELREIPDGEFERRVDWHRRHPDVRVYPTTHTGIARPLADHDERVQLAVI 240
DB      181 ILALGSRPEELREIPDGEFERRVDWHRRHPDVRVYPTTHTGIARPLADHDERVQLAVI 240

OY      241 GGGAGQLARLVGSGHPVFRHAECSVLVRR 272
DB      241 GGGAGQLARLVGSGHPVFRHAECSVLVRR 272

RESULT 2
C70942      hypothetical protein Rv2028c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
```

C/Accession: C70942  
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
C/Accession: C70942  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-279 <COL>  
A:Cross-references: UNIPROT:O53474; UNIPARC:UPI00000D5EDC; GB:AL021899; GB:AL123456; NID:  
A:Experimental source: strain H37Rv  
C/Genetics:  
A:Gene: Rv2028c

Query Match 29.8%; Score 416; DB 2; Length 279;  
Best Local Similarity 36.8%; Pred. No. 2e-24;  
Matches 98; Conservative 46; Mismatches 116; Indels 6; Gaps 2;

11 TIIYIGDSHAATTAALMGVDEAISRRAVRLVSVIKPTSPDDYD---RDLAAERSL 67  
9 SIIVGIDGSKRAVQALMAVDEASRDIPRLVIAIEPDDGYAAGAAKLAEMAV 68  
68 REAGAAVEAAGKLVKIEITDIPRGAPVLAASRDAMIYGVSGVIGRVAASIIIGSTATE 127  
69 RYAFVAADRAAPVKEVEITQERPVTSILIRASAAALVCGAIGVHHFRPERVSTAA 128  
128 LAERAKHCVAVMRKRVDPADINDIIVRMTPADPNEAVLEAAREALROAPILALGCR 187  
129 LALAAOCVAVLRPHRVPIGRDAWIVVEADGSSDIGVLLGAVAAEARLDSFVRVYVCR 188  
188 PEELREIPD---GEFERVODMHHRPDVRVYPIITHTGIAFLADHDERVOLAVIGGE 244  
189 QSGVGTGDDVRAISIDRLAARMPRYPRVQSAAVHELDYLAGLGRSVHMYVLSASD 248  
245 AGQLARLVGPSGHPVFRHAECSTLVV 270  
249 QEHVEQLVGAAPGNVLAQAGCTLLVV 274

RESULT 3  
G70645  
hypothetical protein Rv3134c - Mycobacterium tuberculosis (strain H37Rv)  
C/Species: Mycobacterium tuberculosis  
C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
C/Accession: G70645  
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
C/Accession: G70645  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-268 <COL>  
A:Cross-references: UNIPROT:P95192; UNIPARC:UPI00000D603B; GB:Z83867; GB:AL123456; NID:  
A:Experimental source: strain H37Rv  
C/Genetics:  
A:Gene: Rv3134c

Query Match 25.6%; Score 357.5; DB 2; Length 268;  
Best Local Similarity 36.5%; Pred. No. 5.5e-20;  
Matches 100; Conservative 35; Mismatches 110; Indels 29; Gaps 6;

10 KTIIVGIDGSHAATTAALMGVDEAISRRAVRLVSVIKPTSPDDYDRDLAAERSLRE 69  
8 RAVVVGIDGSRALAAHMAVDEAVNRDIPRLVYVIDPSQLSAAGCGGSAARALAH 67  
70 AQAFAAAGKLVKIEITDIPRGAPVLAASRDAMIYGVSGVIGRVAASIIIGSTATELA 129

68 ASRKEVETGDPVKIETVLGCRPLTKMQSSRSAAMLCVSGVD-HVGRGRGVAATTLA 126  
130 EKACHPVAVMRKRVDPADINDIIVRMTPADPNEAVLEAAREALROAPILALGCR 187  
127 GSAICPFAVIVHPSPAEATTSQVSAVAEV-----DNQVLRHAFEEARLGLVPLRAVA-- 180  
188 PEELREIPD-----EFERRVODMHHRPDVRVYPIITHTGIAFLADHDERVOL 237  
181 -VHAETPPDVEQGSRLAHVLSRLAHMTRLYPEVVDRAIAGSACRHLAANAQPGQL 239  
238 AVTGGGAGQLARLVGPSGHPVFRHAECSTLVV 271  
240 FVADSHSAHELCGAYOPG-----CAVLTVR 264

RESULT 4  
A70942  
hypothetical protein Rv2026c - Mycobacterium tuberculosis (strain H37Rv)  
C/Species: Mycobacterium tuberculosis  
C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
C/Accession: A70942  
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
C/Accession: A70942  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-294 <COL>  
A:Cross-references: UNIPROT:O53472; UNIPARC:UPI00000D6040; GB:AL021899; GB:AL123456; NID:  
A:Experimental source: strain H37Rv  
C/Genetics:  
A:Gene: Rv2026c

Query Match 21.4%; Score 298.5; DB 2; Length 294;  
Best Local Similarity 28.8%; Pred. No. 1.9e-15;  
Matches 83; Conservative 53; Mismatches 119; Indels 33; Gaps 8;

12 TIIYIGDSHAATTAALMGVDEAISRRAVRLVSVIKPT-----HSPDDYDRDLAH 62  
10 ILVGVDSQAQSNAAVMAAREAVMRQLPITLIVAVVVGWVGQLYAMTMOXO--N 67  
63 ABRSLRAQAGAVE---AAGLVKIEITDIPRGAPVLAASRDAMIYGVSGVIGRVAAS 119  
68 AQQVTEQAREALTNLSGESKRPQVHTLVSNVPTLIDASQAAMLVVSGGKALGR 127  
120 ILGSTATLEAKHCPVAVMRKRVDPADINDIIVRMTPADPNEAVLEAAREALRO 178  
128 ILGSGISALHHRCPVAIIHSGNATPPSDAP-VLVGIDGSPASEAATLALADEASRR 186  
179 A-----PILALGSRPEELREIPDGEFERVODMHHRPDVRVYPIITHTGT 224  
187 VDLVLAHMTDGLMFPLGMDWRERERKEBA--VLAERLGMQBOYPDVVRHSLVCDKP 244  
225 ARFLADHDERVOLAVIGG-GEAGQLARLVGPSGHPVFRHAECSTLVV 271  
245 ARWLHESBOQLVVVSGHGRGFGSMLLGVSASAVHVRIVYIVR 292

RESULT 5  
C70759  
hypothetical protein Rv2005c - Mycobacterium tuberculosis (strain H37Rv)  
C/Species: Mycobacterium tuberculosis  
C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
C/Accession: C70759  
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: F70758  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-295 <COL>  
A:Cross-references: UNIPROT:Q10851; UNIPARC:UPI000013B62F; GB:Z74025; GB:AL123456; NID:G  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV2005c

Query Match 20.5%; Score 286; DB 2; Length 295;  
Best Local Similarity 29.6%; Pred. No. 1.7e-14;  
Matches 85; Conservative 48; Mismatches 126; Indels 28; Gaps 8;

Qy 12 IIVGIDGSHAITAALWGVDEAISRVPRLVSYIKP-----PTHPSPDDY-----DPDL 60  
Db 10 VVVGVDGSLSDPAACWGCATDAAMNIPDLTVHVNADVATWPPMPPEYETGWQEDGCR 69  
Qy 61 AAHRSLSRBAQSAVEAAGKLVKIEDIPRGAPVLYEASRDAMI CVSGVIGRYASSI 120  
Db 70 QIVANAVLAKEAAGADKKL-SVNSLVTSTPVPYPMVEISNEAEVNVLGSSGRGLAKGL 128  
Qy 121 LGSTATLAEKACHPVAWRSKVDQPASDINW--IVVMTDAPDNEAVLYEAAREAKLRQ 178  
Db 129 LGSVSSSLVRRAGCGVAVIYHSD-DAVIDPQHPAPLVGIDGSPVSELATAVAFDEASRG 187  
Qy 179 APIIALGGRPE-ELREIPDGEPE-----RRVDDMHRRPDVRYPTTHGTGAR 226  
Db 188 VELAVHAMSVEVEVELPGLDFSAVQDEALSLAERLAKGQERYPDVEFSRVVVCDRPAR 247  
Qy 227 FLADHDERVOLAVIGG-GEAGOLARLVGSGHPVFRHAECSVLVVR 272  
Db 248 KLVQKASASQLVVVGSHRGGLTGMLDGSVSNVLAHARPVYVARQ 294

## RESULT 6

F70572  
hypothetical protein RV2623 - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
C:Accession: F70572  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;  
Rajandream, M.A.; Rogers, J.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: F70572  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-297 <COL>  
A:Cross-references: UNIPROT:O06189; UNIPARC:UPI000003326C; GB:Z95387; GB:AL123456; NID:G  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV2623

Query Match 19.1%; Score 267; DB 2; Length 297;  
Best Local Similarity 29.3%; Pred. No. 4.9e-13;  
Matches 84; Conservative 42; Mismatches 131; Indels 30; Gaps 8;

Qy 12 IIVGIDGSHAITAALWGVDEAISRVPRLVSYIKP-----THPSP-----DYYDR 58  
Db 10 IIVGIDGSHAITAALWGVDEAISRVPRLVSYIKP-----THPSP-----DYYDR 58  
Qy 59 DLAHRSLSRBAQSAVEAAGKLVKIEDIPRGAPVLYEASRDAMI CVSGVIGRYASSI 118  
Db 70 HL--IDDAKLVVEQSLRAGP-PTVHSEIVPAAAPFLLVDSKQVLAIVVGGCLSGRRPG 126  
Qy 119 SILGSTATLAEKACHPVAWRSKVDQPASDINWIVVMTDAPDNEAVLYEAAREAKLR 177

Db 127 RLIGSVSSGLIRHACPVAVIIHDEDSVMPHPQAPVLVGDGSSASSELATIAFDEASRR 186  
Qy 178 QAPIALGGRPE-ELREIP-----DGEFEREVQDWHRRPDPVRYPTTHGTGA 225  
Db 187 NVDVLAHMSDVVSEMPGIDWPAATQSMAEQVLAERLAGQERYPVAVIRVVRQPA 246  
Qy 226 FLADHDERVOLAVIGG-GEAGOLARLVGSGHPVFRHAECSVLVVR 271  
Db 247 RQLVGRBEAQVLVVGSRGGRGVAGMLVGSVGETVAQLARPVIVAR 293

## RESULT 7

B70758

hypothetical protein RV1996 - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
C:Accession: B70758  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;  
Rajandream, M.A.; Rogers, J.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: B70758  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-317 <COL>  
A:Cross-references: UNIPROT:Q10862; UNIPARC:UPI000013B4FB; GB:Z74025; GB:AL123456; NID:G  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV1996

Query Match 18.5%; Score 258.5; DB 2; Length 317;  
Best Local Similarity 28.3%; Pred. No. 2.4e-12;  
Matches 85; Conservative 47; Mismatches 127; Indels 41; Gaps 9;

Qy 12 IIVGIDGSHAITAALWGVDEAISRVPRLVSYIKPTHPSPD-----DYYDR----- 59  
Db 10 IIVGVDSPCHSTAVENBARDAQMRVNAIVVQVPPVITTAPEGMAREYSRFGSAQKREI 69  
Qy 60 -----LAHRSLSRBAQSAVEA-AGKLVKIEDIPRGAPVLYEASRDAMI CVGS 110  
Db 70 VEHSYLVAAQHQIVEQAHKVLAELASSGRAAQITGEVLAHQIVPTLANISQVAMVVLGY 129  
Qy 111 VGIRYVASSILGSTATLAEKACHPVAWRSKVDQPASDINW-IVVMTDAPDNEAVLEY 169  
Db 130 RGGQAVAGALIGSVSSSLVRRHAGPVAVIPEE-PRPARPPHAPVVGIDGSPVSGLAERI 188  
Qy 170 AAREAKLRQAPVLA-----LGGPPELEIPDGE---PERRVQDWHRRHPD 212  
Db 189 AFDEASRRGVDVLAHMSDQGLDFPRLMVAPIEMRNLEDBQKMLARKLSGQDXYPD 248  
Qy 213 VRVYPITTHGTARFLADHDERVOLAVIGG-GEAGOLARLVGSGHPVFRHAECSVLVVR 271  
Db 249 VVHAKVVVCRPARRLLELAQTQLVVGSHRGGRFGGMHIGSVRAVNVNSGAPVIVAR 308

## RESULT 8

T37029

hypothetical protein SCU12.10c - Streptomyces coelicolor

C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T37029  
R:Murphy, L.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, August 1999  
A:Reference number: 221619  
A:Accession: T37029  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-288 <MOR>  
A:Cross-references: UNIPROT:Q9RI48; UNIPARC:UPI000000DB3B4; EMBL:AL109989; PTDN:CAB53422..  
A:Experimental source: strain A3 (2)



C/Genetics:  
A:Gene: SCOEDB:SCJ12.10C

Query Match 17.4%; Score 243.5; DB 2; Length 288;  
Best Local Similarity 26.8%; Pred. No. 2.9e-11;  
Matches 85; Conservative 36; Mismatches 123; Indels 51; Gaps 9;

QY 12 IIVGIDGSHAATTAALMGVDEAISRVAFLRVSVI-----KPT-HPSPDD 55  
DB 5 MVVGVDSGESSLGAIVDMAADEALHEVPLRIIVHVRMDRYEGASLARELGPSGHVTTDD 64  
QY 56 YDRDLAHERSLREASQAVEAAGLVKLETDPICPGPAGPVLYEASRDMEICVSGVIGR 115  
DB 65 I---LAVATRRARRHHPDLAVTTATAEPEY-----VLLREARNASAVILGTRGGE 114  
QY 116 YASSILGSTATLEAKAPVAVMRSKVDOPASD---INWIVVMTDPADEAVLEYAR 172  
DB 115 LAGLILGSVSLTVATMSDCPVVTRGDHDDRAAGGRGRIVGVADAP--TAAVRPCE 172  
QY 173 EAKLRQAPILA-----LGGRPEELREIPDG-EFERRVODMHHRRPDRV 215  
DB 173 EARRGALADVAMRCPTHDPVHPLAGTPERHERAKALEALAD---APADVRL 229  
QY 216 YPTTHTGIAFLADHDERVOLAVIIGGEAGQLARLVGPSGHVFRRAECGVLYV 270  
DB 230 RRRTAEGSGRVLSSASHADLLVGRRRPGCFGRILGRVAHTLLHSAECVAVV 284

## RESULT 9

hypothetical protein SCJ1.29C - Streptomyces coelicolor  
C/Species: Streptomyces coelicolor  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: T36962  
R:Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, September 1999  
A/Reference number: 221607  
A/Accession: T36962  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-263 <SEB>  
A/Cross-references: UNIPROT:Q9R1Y5; UNIPARC:UPI00000DB39B; EMBL:AL109962; PIDN:CAB53147.  
A/Experimental source: strain A3(2)  
C/Genetics:  
A:Gene: SCOEDB:SCJ1.29C

Query Match 16.5%; Score 230; DB 2; Length 283;  
Best Local Similarity 30.0%; Pred. No. 3e-10;  
Matches 87; Conservative 35; Mismatches 122; Indels 46; Gaps 9;

QY 12 IIVGIDGSHAATTAALMGVDEAISRVAFLRV-----SVIKTPSPDDYDRD 59  
DB 5 LVVGVDSGSEPLRAVDMADEBALAHVPLVWVFGDLMRYEGALALARPGRPSIDMQADD 64  
QY 60 LAHAERSLREASQAVEAAGK--LVKLETDPICPGPAGPVLYEASRDMEICVSGVIGRY 116  
DB 65 IL-----AAAIRAGRRHHPDLV-VTTEVPDEABEHALLCAGRNASMIWGSRGSGI 115  
QY 117 ASSILGSTATLEAKAPVAVMRSKVDOPR--SPINWIVVMTDPADEAVLEYARE 174  
DB 116 ADRLIGSVSRVVAAGSDCPVVVLRGNHNDRAIGGRNRIIVGVGEV--SASVLRLEFMEA 173  
QY 175 KLARQAPILAGRPBELRE-----IPDGFERRVODMHHRR-----PDVRYVPTT 220  
DB 174 RLROVPAAVAAMCRPHETIDHPLPAGSPRRRYERAAKELALBEDAPDVAVRKYTV 233  
QY 221 HTGIARFLADHDERVOLAVIIGGEAGQLARLVGPSGHVFRRAECGVLYV 270  
DB 234 EGPARAVLPAASABAGLVIGRRQGRIGRV-----AAHVLRSACPVVVV 279

## RESULT 10

T37031

hypothetical protein SCJ12.12C - Streptomyces coelicolor

C/Species: Streptomyces coelicolor  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: T37031  
R:Murphy, L.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, August 1999

QY 12 IIVGIDGSHAATTAALMGVDEAISRVAFLRVSVIKPTHPSPDDYDRDLAHA----- 63  
DB 3 RTIVGIDGSPESRAAEMARREGTLRRVPVRLHWQPV-PPD-----MAQAPILGAE 55  
QY 64 -----ERSLREASQAVEAAGLVKLETDPICPGPAGPVLYEASRDMEICVSGVIGRY 117  
DB 56 THQWTERIPDPAEAGLRHNPGEVETTEQATGNPADALAGTLDDELVLGSRALSGLT 115  
QY 118 SSILGSTATLEAKAHCPVAVNR-----SKVDOPASDINWIVVMTDPADE 164  
DB 116 GFLVGSYGOSVIATETPTVLVRAAGADENHLKPTGISAAITGRPVVVGGLDTSPE 175

## RESULT 11

Query Match 16.1%; Score 224.5; DB 2; Length 301;  
Best Local Similarity 25.4%; Pred. No. 8.6e-10;  
Matches 78; Conservative 42; Mismatches 130; Indels 57; Gaps 8;

QY 10 KTIIVGIDGSHAATTAALMGVDEAISRVAFLRVSVIKPTHPSPDDYDRDLAHA----- 63  
DB 3 RTIVGIDGSPESRAAEMARREGTLRRVPVRLHWQPV-PPD-----MAQAPILGAE 55  
QY 64 -----ERSLREASQAVEAAGLVKLETDPICPGPAGPVLYEASRDMEICVSGVIGRY 117  
DB 56 THQWTERIPDPAEAGLRHNPGEVETTEQATGNPADALAGTLDDELVLGSRALSGLT 115  
QY 118 SSILGSTATLEAKAHCPVAVNR-----SKVDOPASDINWIVVMTDPADE 164  
DB 116 GFLVGSYGOSVIATETPTVLVRAAGADENHLKPTGISAAITGRPVVVGGLDTSPE 175  
QY 165 AVLEYAREAKLRQAPI-----LALGSR-EELREIPDGFERRVODMHN 208  
DB 176 AVLSFAEEARRRRAPLTVARAMNLPSYTYSLAAGDPREBELARQAALGALLPWRE 235  
QY 209 RHPDVRVYPTTHTGIRFLADHDERVOLAVIIGGEAGQLARL-----VPSGHPFRHA 263  
DB 236 KYPDVEETETCRLGSPAEHLIDARDAVLVVG---RRIRSPFGVHIGAVAAVHNA 291  
QY 264 ECGSVLYV 270  
DB 292 TTPVAVV 298

## RESULT 11

conserved hypothetical protein SCJ1.16C - Streptomyces coelicolor  
C/Species: Streptomyces coelicolor  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: T36949  
R:Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, September 1999  
A/Reference number: 221607  
A/Accession: T36949  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-294 <SEB>  
A/Cross-references: UNIPROT:Q9R1Z8; UNIPARC:UPI00000DB38E; EMBL:AL109962; PIDN:CAB53134.3  
A/Experimental source: strain A3(2)  
C/Genetics:  
A:Gene: SCOEDB:SCJ1.16C

Query Match 15.7%; Score 220; DB 2; Length 294;  
Best Local Similarity 24.9%; Pred. No. 1.8e-09;  
Matches 73; Conservative 43; Mismatches 133; Indels 44; Gaps 6;

QY 12 IIVGIDGSHAATTAALMGVDEAISRVAFLRV-----SVIKTPHPSPDDYDRD 59  
DB 8 LVVGVDSGDSLALIDVADEAQRGRLRLVYASLWERYEGALPAMGRSPS-----E 61  
QY 60 LAHAERSLREASQAVEAAGLVKLETDPICPGPAGPVLYEASRDMEICVSGVIGRYASS 119

Db 62 QVMAENIVGTAERVRRYDPLGTTDTDTVPAPAAVSAALAEGRHATAVVTSGRGCELKGA 121  
 Qy 120 ILGSTATELAKKACPVAVMRSKYDQSPASDINMIVRMTPADPNEAVLEAAREAKLRQA 179  
 Db 122 LLGVSISLAVARADCPVAVVVGKSKALSGSHERVILGAGDPDTSGAAVRPAFPAFREADVAGC 181  
 Qy 180 PILAL-----GGRPEELREIPDGEFERR-----VQDWHHRHPDVRYVPI 218  
 Db 182 ELDVVRAMRCPAYENADEGASDSDSDP-----ERRASALIDTVAEAAAEHPSVRLAKT 237  
 Qy 219 TTHGIAFIADHDHDERVOLAVIYG-GEAGQALRVGSGHVFERRAESVLYV 270  
 Db 238 TIEGPARKVLVHRTAAADLVVVGARHRSRGHGLQGRVTHTLLOHAACPVAHV 290

RESULT 12  
 T36963  
 hypothetical protein SCJ1.30c - Streptomyces coelicolor  
 C/Species: Streptomyces coelicolor  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C/Accession: T36963  
 R/Seeger: K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, September 1999  
 A/Reference number: Z21607  
 A/Accession: T36963  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-328 <SER>  
 A/Cross-reference: UNIPROT:Q9RIY4; UNIPARC:UPI00000DB39C; EMBL:AL109962; PIDD:CAM53148.  
 A/Experimental source: strain A3(2)  
 A/Genetics:  
 A/Gene: SCODB:SCJ1.30c

Query Match 13.3%; Score 186; DB 2; Length 328;  
 Best Local Similarity 26.8%; Pred. No. 8, 2e-07;  
 Matches 82; Conservative 38; Mismatches 130; Indels 56; Gaps 12;  
 Qy 10 KTIIVGIDGSHAATLALMGVDEAISRAVPLRLVSVIKPTSPDPDYRDIA-----HA 63  
 Db 31 RTIVGIDGSPESHAAEMARAEALRLPRLHVMFPA-PAYALQDSITGAKTHQHW 89  
 Qy 64 ERSIREAGSAVEAAGKLVKLTETDIPRGAPVLYEASDAEMICVGS--VGIGRYASSI 120  
 Db 90 ERVPQKVEGRLRLHPGVVVTSDQSGPADTLVADAGALVYGSAPGLGCF--L 146  
 Qy 121 LGSATTELAKKACPVAVMRS-----KYDQ-----PASDINMIVRMTP--APDNRAV 166  
 Db 147 AGSVGQSVIAHSETPVVLVRAGEQAAGEHVDSVGVPPAANRRRRPVVGLDIGSPD-DGV 205  
 Qy 167 LEVAREAKLRQA-----PILALGGRPEELREIPDGEFERRVOD-----205  
 Db 206 LSPFDEAGRRGTAVHVVGKRRPPRYPSLIRAGVPDQ-----GVARRKADLRLLLP 259  
 Qy 206 WNHHRPDVRYVPIITHTGIAFIADHDHDERVOLAVIYG-GEAGQALRVGSGHVFERRAE 264  
 Db 260 WRQGSVPDVEVEAASRPSPADLLEASHEASLVVVGRIIRPSLGVNHGAVAHVLRVVS 319  
 Qy 265 CSVLVY 270  
 Db 320 APVAHV 325

RESULT 13  
 T37036  
 hypothetical protein - Streptomyces coelicolor  
 C/Species: Streptomyces coelicolor  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C/Accession: T37036  
 R/Murphy, L.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, August 1999  
 A/Reference number: Z21619  
 A/Accession: T37036  
 A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA  
 A/Residues: 1-201 <MUR>  
 A/Cross-reference: UNIPROT:Q9RI41; UNIPARC:UPI00000DB3EB; EMBL:AL109969; PIDD:CAM53431.  
 A/Experimental source: strain A3(2)  
 C/Genetics:  
 A/Gene: SCODB:SCJ12.19c

Query Match 11.5%; Score 160.5; DB 2; Length 201;  
 Best Local Similarity 31.9%; Pred. No. 3, 9e-05;  
 Matches 60; Conservative 30; Mismatches 73; Indels 25; Gaps 7;  
 Qy 12 IIVGIDGSHAATLALMGVDEAISRAVPLRLVSVIKPTSPDPDYR-----DLAHERS 66  
 Db 5 IAVGIDSPASLAAMHAAHEARRRGSGITLVHWHRRARPFP--YLRVGSTERAMART 62  
 Qy 67 LREAQSAVEAAGKLVKLTETDIPRGAPVLYEASDAEMICVGSVIGRYASSIIGSTAT 126  
 Db 63 LEEAVRSVRSNHPGLRTTERLVCDATVYALVTAADEMLVLSFGFGPGVPGFTGVSQ 122  
 Qy 127 ELAEKACPVAVM--RSKYDQ--PASD-----INMIVRMTPADPNEAVLEY 169  
 Db 123 RVVARADHPVLYVRAAGSADBEHLPAADGLAPEIPETPYRDVLYGRVDRPCDE-VLDF 181  
 Qy 170 AAREAKLR 177  
 Db 182 AFEAARRR 189

RESULT 14  
 T36954  
 hypothetical protein SCJ1.21 - Streptomyces coelicolor  
 C/Species: Streptomyces coelicolor  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C/Accession: T36954  
 R/Seeger: K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, September 1999  
 A/Reference number: Z21607  
 A/Accession: T36954  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-152 <SER>  
 A/Cross-reference: UNIPROT:Q9RI23; UNIPARC:UPI00000DB393; EMBL:AL109962; PIDD:CAM53139..  
 A/Experimental source: strain A3(2)  
 C/Genetics:  
 A/Gene: SCODB:SCJ1.21  
 C/Superfamily: Escherichia coli ybdQ protein

Query Match 10.7%; Score 149; DB 2; Length 152;  
 Best Local Similarity 29.3%; Pred. No. 0.0002;  
 Matches 44; Conservative 26; Mismatches 62; Indels 18; Gaps 5;  
 Qy 6 EPTKTIIVGIDGSHAATLALMGVDEAISRAVPLRLVSVIKPTH-----PS-----PD 54  
 Db 5 EPTAR-VVVGIDGSPSSYALRNA--DRYAAVG---GVAEAVHWDTSPAVGAPAPAI 57  
 Qy 55 DYEDDLAARSLREAQSAVEAAGKLVKLTETDIPRGAPVLYEASDAEMICVGSVIG 114  
 Db 58 DPDFDLQAKERRPAALAEATPPGRRPGLKEILVSGPSETLIRASGALVVGRRGRG 117  
 Qy 115 RYASSIIGSTATLEAKKACPVAVMRSKVD 144  
 Db 118 AFARMLGVSQRCAQHAACPVVVROETE 147

RESULT 15  
 C86231  
 hypothetical protein [imported] - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
 C/Accession: C86231  
 R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hultzer, L.





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 04:56:43 ; Search time 69.5674 Seconds  
(without alignments)  
2758.529 Million cell updates/sec

Title: US-10-617-038-27

Perfect score: 1397  
Sequence: 1 MSGRGEPTMTITVIGIDGSH.....GPSGHPVRHAECSVLVVR 272

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot\_05.80:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1397	100.0	272	Q7TY74_MYCBO	Q7EY74 mycobacteri
2	1397	100.0	272	O06188_MYCTU	O06188 mycobacteri
3	416	29.8	279	Q7TZ44_MYCBO	Q7E444 mycobacteri
4	416	29.8	279	O53474_MYCTU	O53474 mycobacteri
5	357.5	25.6	268	Q7TX68_MYCBO	Q7K668 mycobacteri
6	357.5	25.6	268	P95192_MYCTU	P95192 mycobacteri
7	342.5	24.5	252	Q73V37_MYCPA	Q73V37 mycobacteri
8	298.5	21.4	294	Q7TZ45_MYCBO	Q7E445 mycobacteri
9	298.5	21.4	294	O53472_MYCTU	O53472 mycobacteri
10	286	20.5	295	Y2005_MYCTU	P64921 mycobacteri
11	286	20.5	295	Y2028_MYCBO	P64922 mycobacteri
12	277.5	19.9	318	O081R9_COREF	O081R9 corynebacte
13	276.5	19.8	293	Q73Z61_MYCPA	Q73Z61 mycobacteri
14	271.5	19.4	294	Q73Z62_MYCPA	Q73Z62 mycobacteri
15	267	19.1	297	Q7TY75_MYCBO	Q7EY75 mycobacteri
16	267	19.1	297	O06189_MYCTU	O06189 mycobacteri
17	266	19.0	301	O081J9_CORGL	O081J9 corynebacte
18	266	19.0	345	O5YV44_MYCFA	O5YV44 nocardia fa
19	260	18.6	294	Q73Z49_MYCPA	Q73Z49 mycobacteri
20	258.5	18.5	317	Y1996_MYCTU	P045F7 mycobacteri
21	258.5	18.5	317	Y2019_MYCBO	P045F8 mycobacteri
22	254	18.2	296	O5YVE7_MYCFA	O5YVE7 nocardia fa
23	253.5	18.1	296	O5YTS4_MYCFA	O5YTS4 nocardia fa
24	249	17.8	300	O08NEU3_CONDI	O08NEU3 corynebacte
25	243.5	17.4	288	O09R148_STRCO	O09R148 streptomyc
26	242.5	17.4	290	O082NM7_STRAM	O082NM7 streptomyc
27	234	16.8	293	O5YVP7_NOCFA	O5YVP7 nocardia fa
28	231	16.5	303	O4JY55_CORJX	O4JY55 corynebacte
29	230	16.5	283	O09R1Y5_STRCO	O09R1Y5 streptomyc
30	227.5	16.3	300	O09K4H8_STRCO	O09K4H8 streptomyc
31	224.5	16.1	301	O09R146_STRCO	O09R146 streptomyc

32	224	16.0	281	2	Q7WZ49_9ACTO	Q7WZ49 nonomuraea
33	220	15.7	294	2	Q0R1Z8_STRCO	Q0R1Z8 streptomyc
34	210	15.0	295	2	O082NK3_STRAM	O082NK3 streptomyc
35	210	15.0	345	2	O083U1_BIFLO	O083U1 bifidobacte
36	205	14.7	312	2	O0K4L5_STRCO	O0K4L5 streptomyc
37	202.5	14.5	345	2	O7UJW6_RHOBA	O7UJW6 rhodospirill
38	202	14.5	343	2	Q4NL54_9MICC	Q4NL54 arthrobacte
39	191	13.7	144	2	Q4NDP1_9MICC	Q4NDP1 arthrobacte
40	186	13.3	328	2	O09R1Y4_STRCO	O09R1Y4 streptomyc
41	180.5	12.9	315	2	O7UER2_RHOBA	O7UER2 rhodospirill
42	169.5	12.1	289	2	Q4NE71_9MICC	Q4NE71 arthrobacte
43	168.5	12.1	151	2	Q4NGW3_9MICC	Q4NGW3 arthrobacte
44	162.5	11.6	155	2	Q4NL67_9MICC	Q4NL67 arthrobacte
45	160.5	11.5	201	2	O09R141_STRCO	O09R141 streptomyc

## ALIGNMENTS

RESULT 1									
ID	Q7TY74_MYCBO	PRELIMINARY;	PRT;	272	AA.				
AC	Q7TY74:								
DT	01-OCT-2003 (TREMBLrel. 25, Created)								
DT	01-OCT-2003 (TREMBLrel. 25, Last sequence update)								
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)								
DE	Hypothetical protein MD2657c.								
GN	OrderedlocusNames=MD2657c;								
OS	Mycobacterium bovis.								
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;								
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium;								
OC	Mycobacterium tuberculosis complex.								
OX	NCBI_TaxID=1765;								
RN	[1]								
RP	NUCLEOTIDE SEQUENCE.								
RC	STRAIN=AF2122/97;								
RX	MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;								
RA	Garnier T., Eglmeier K., Camus J.-C., Medina N., Mansoor H.,								
RA	Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,								
RA	Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,								
RA	Parthill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,								
RT	"The complete genome sequence of Mycobacterium bovis.";								
RL	Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).								
DR	EMBL; BX248343; CAD94842.1; -; Genomic DNA.								
DR	GO; GO:0006950; P:response to stress; IEA.								
DR	InterPro; IPR006015; Ugp.								
DR	InterPro; IPR006016; UspA.								
DR	Pfam; PF00582; Usp, 2.								
DR	PRINTS; PR01438; UNVRS1STRESS.								
KW	Complete proteome.								
SQ	SEQUENCE 272 AA; 29400 MW; 3EE3A17B3C5720DB CRC64;								
Query Match									
		100.0%;	Score 1397;	DB 2;	Length 272;				
		Best Local Similarity	100.0%;	Pred. No. 2.1e-99;					
		Matches 272;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;			
QY	1	MSGRGEPTMTITVIGIDGSHAITAALMGVDEAISRAVPLRLTVIKETHPSPDDYDDL	60						
DB	1	MSGRGEPTMTITVIGIDGSHAITAALMGVDEAISRAVPLRLTVIKETHPSPDDYDDL	60						
QY	61	AHAERSLREAQSAVEAAGKLVKIEFTDIPRGPAFVLVEASRDAMI CVSGVIGRYASSI	120						
DB	61	AHAERSLREAQSAVEAAGKLVKIEFTDIPRGPAFVLVEASRDAMI CVSGVIGRYASSI	120						
QY	121	LGSTATLAEAKACGPAVMRSKVDQPSADINMIVRMTPDADNDAVLEYAREKLRAP	180						
DB	121	LGSTATLAEAKACGPAVMRSKVDQPSADINMIVRMTPDADNDAVLEYAREKLRAP	180						
QY	181	ILATGGRREERREIPDGEFERRVODMHRHHPDVAVYPTTTGTGIAFLADDERVQLAVI	240						
DB	181	ILATGGRREERREIPDGEFERRVODMHRHHPDVAVYPTTTGTGIAFLADDERVQLAVI	240						
QY	241	GGGAGQLARLVGSGHPVFRHAECSVLVVR	272						

Db 241 GGGAGQIARLVGSPGHVFRHACSVLVVR 272

RESULT 2

ID 006188\_MYCTU PRELIMINARY; PRT; 272 AA.  
AC 006188; Q7D6V6;  
DT 01-JUL-1997 (TReMBLrel. 04, Created)  
DT 01-JUL-1997 (TReMBLrel. 04, last sequence update)  
DT 13-SEP-2005 (TReMBLrel. 31, last annotation update)  
DE Hypothetical protein (Universal stress protein family).  
GN OrderedLocusNames=MT2699, RV2624c;  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;  
OC Mycobacterium tuberculosis complex.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230; DOI=10.1038/21159;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C.M.,  
RA Harris D.E., Gordon S.V., Eigmeier K., Gas S., Barry C.E. III,  
RA Tekela F., Badcock K., Basham D., Brown D., Chillingworth T.,  
RA Connor R., Davies R.M., Devlin K., Feltwell T., Gentles S., Hamlin N.,  
RA Holroyd S., Hornsby T., Jagels K., Krogh J.A., McLean J., Moule S.,  
RA Murphy J.D., Oliver S., Osborne J., Quail M.A., Ratnam M.A.,  
RA Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.,  
RT "Defining the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544 (1998).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=CDC 1551 / Oshkosh;  
RX MEDLINE=22206494; PubMed=12218036;  
RX DOI=10.1128/JB.184.19.5479-5490.2002;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J.D., DeBoy R.T., Dodson R.J., Gwinn M.L., Haft D.H.,  
RA Hickey E.K., Kolonay J.F., Nelson W.C., Umayam L.A., Esmailova M.D.,  
RA Salzberg S.L., Delcher A., Uitterlinden T.R., Weidman J.F., Kouri H.M.,  
RA Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,  
RA Fraser C.M.,  
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains."  
RL J. Bacteriol. 184:5479-5490 (2002).  
RX EMBL, BX842580; CAB08618.1; -; Genomic\_DNA.  
RX EMBL, AE000516; AKA47015.1; -; Genomic\_DNA.  
DR PIR, G70572; G70572.  
DR TIGR, MT2699; -.  
DR TubercuList; RV2624c; -.  
GO GO:0006950; P:response to stress; IEA.  
DR InterPro; IPR006015; Unp.  
DR InterPro; IPR006016; Unp.  
DR Pfam; PF00582; Unp; 2.  
DR PRINTS; PR01438; UNRS1STRESS.  
KW Complete proteome.  
SQ SEQUENCE 272 AA; 29400 MW; 3EE3A17B3C5720DB CRC64;

Query Match 100.0%; Score 1397; DB 2; Length 272;  
Best Local Similarity 100.0%; Pred. No. 2.1e-99;  
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSGGEPTMTTIVIGIDGSHAITTAALMGVDEAISRVPRLVSVIKPTHTSPDDYDDL 60  
DB 1 MSGGEPTMTTIVIGIDGSHAITTAALMGVDEAISRVPRLVSVIKPTHTSPDDYDDL 60  
QY 61 AAERSLAEAGSAVAAGKLVKIFTDIPRGAPVPLVEASRDAMICVSGVIGYASSI 120  
DB 61 AAERSLAEAGSAVAAGKLVKIFTDIPRGAPVPLVEASRDAMICVSGVIGYASSI 120  
QY 121 LGSTATLEAKAHCPVAVMRSKVDOPASDINMIVVMTDAPDNEAVLEVAAREAKLRQAP 180

Db 121 LGSTATLEAKAHCPVAVMRSKVDOPASDINMIVVMTDAPDNEAVLEVAAREAKLRQAP 180  
QY 181 ILALGRRPEELREIPDGEFFERRVODMHRHDPVAVVYPTTHTGTARFLADHDERVOLAVI 240  
DB 181 ILALGRRPEELREIPDGEFFERRVODMHRHDPVAVVYPTTHTGTARFLADHDERVOLAVI 240  
QY 241 GGGAGQIARLVGSPGHVFRHACSVLVVR 272  
DB 241 GGGAGQIARLVGSPGHVFRHACSVLVVR 272

RESULT 3

ID 07T244\_MYCBO PRELIMINARY; PRT; 279 AA.  
AC 07T244;  
DT 01-OCT-2003 (TReMBLrel. 25, Created)  
DT 01-OCT-2003 (TReMBLrel. 25, last sequence update)  
DT 01-MAR-2004 (TReMBLrel. 26, last annotation update)  
DE Hypothetical protein Mb2053c.  
GN OrderedLocusNames=Mb2053c;  
OS Mycobacterium bovis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;  
OC Mycobacterium tuberculosis complex.  
OX NCBI\_TaxID=1765;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=AF123/97;  
RX MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;  
RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,  
RA Pryor M., Dutfoy S., Grondin S., Lacroix C., Monempe C., Simon S.,  
RA Harris B., Alkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.,  
RT "The complete genome sequence of Mycobacterium bovis."  
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).  
RX EMBL, BX248341; CAB96906.1; -; Genomic\_DNA.  
DR GO; GO:0006950; P:response to stress; IEA.  
DR InterPro; IPR006015; Unp.  
DR InterPro; IPR006016; Unp.  
DR Pfam; PF00582; Unp; 1.  
DR PRINTS; PR01438; UNRS1STRESS.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 279 AA; 29441 MW; 52F240F9FE466AC9 CRC64;

Query Match 29.8%; Score 416; DB 2; Length 279;  
Best Local Similarity 36.8%; Pred. No. 7.8e-24;  
Matches 98; Conservative 46; Mismatches 116; Indels 6; Gaps 2;

QY 11 TIIVIGIDGSHAITTAALMGVDEAISRVPRLVSVIKPTHTSPDDYD--RDLAAERSL 67  
DB 9 SIIVIGIDGSKPAVQALMAVDEAASRDIPRLTVAIPEDDPGVAAHGAARKLAAANNAV 68  
QY 68 REAGSAVEAGKLVKIFTDIPRGAPVPLVEASRDAMICVSGVIGYASSIIGSTATE 127  
DB 69 RYAFPAVAADRPKVEVEITQERPTSLIRASAAALVCAGALGVHFRPERVSTPAA 128  
QY 128 LAERAPCPVAVMRSKVDOPASDINMIVVMTDAPDNEAVLEVAAREAKLRQAPILALGCR 187  
DB 129 LALSAQCPVAIVRHRPIGRDAIMYVEADGSDIGVLGAWAERLADSPRVVYTCR 188  
QY 188 PEELREIPD--GEFFERRVODMHRHDPVAVVYPTTHTGTARFLADHDERVOLAVIGGE 244  
DB 189 QSGVGTGDVDVRAISLDRLARWQGRYDVRVQSAAVGELLDYLAGRSGVHVVLSASD 248  
QY 245 AGQLARLVGSPGHVFRHACSVLVV 270  
DB 249 QEHVEQLVGAAGNAVLQEGACTLLV 274

RESULT 4

ID 053474\_MYCTU PRELIMINARY; PRT; 279 AA.

AC 053474; Q7D7L5;  
 DT 01-JUN-1998 (TReMBLrel. 06, Created)  
 DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)  
 DT 13-SEP-2005 (TReMBLrel. 31, Last annotation update)  
 DE Hypothetical protein (Universal stress protein family).  
 GN OrderedLocustNames=MT2087, Rv2028c;  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;  
 OC Mycobacterium tuberculosis complex.  
 NC NCB1\_TaxID=1773;  
 RX NUCLEOTIDE SEQUENCE.  
 RP STRAIN=H37Rv;  
 RC MEDLINE=96295987; PubMed=9634230; DOI=10.1038/31159;  
 RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C.M.,  
 RA Harris D.E., Gordon S.V., Bigliameier K., Gas S., Barry C.E. III,  
 RA Tekle A.F., Badcock K., Basham D., Brown D., Chillingworth T.,  
 RA Connor R., Davies R.M., Devlin K., Felwell T., Gentles S., Hamlin N.,  
 RA Holroyd S., Hornby T., Jagels K., Ketchell A., McLean J., Moule S.,  
 RA Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,  
 RA Rogers J., Rutter S., Seeger K., Skellton S., Squares S., Squares R.,  
 RA Sultson J.E., Taylor K., Whitehead S., Barrall B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RX MEDLINE=22206494; PubMed=12218036;  
 RX DOI=10.1187/BB.184.19.5479-5490.2002;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J.D., DeBoy R.T., Dodson R.C., Gwinn M.L., Haft D.H.,  
 RA Hickey E.K., Kolonay J.F., Nelson W.C., Umayam L.A., Brmlaeva M.D.,  
 RA Salzberg S.L., Delcher A., Utterback T.R., Melman J.F., Khouiri H.M.,  
 RA Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL J. Bacteriol. 184:5479-5490(2002).  
 DR EMBL; BX942578; CA17242.1; --; Genomic\_DNA.  
 DR EMBL; AE000516; AAK46366.1; --; Genomic\_DNA.  
 DR PIR; C70942; C70942.  
 DR TIGR; MT2087; --  
 DR TubercuList; Rv2028c; --  
 DR GO; GO:0006950; P:response to stress; IEA.  
 DR InterPro; IPR006015; Unp.  
 DR InterPro; IPR006016; UnpA.  
 DR Pfam; PF00582; Unp; 1.  
 DR PRINTS; PR01438; UNYRS1STRESS.  
 KW Complete proteome; Hypothetical protein.  
 SO SEQUENCE 279 AA; 29441 MW; 52F240F9FE466AC9 CRC64;

Query Match 29.8%; Score 416; DB 2; Length 279;  
 Best Local Similarity 36.8%; Pred. No. 7.8e-24;  
 Matches 98; Conservative 46; Mismatches 116; Indels 6; Gaps 2;

QY 11 TTIIGTGGSHAAITAAWGVDEAISRVAFLVSVIKPTHPSPDDYD---RDLAHAESL 67  
 DB 9 SIVVGIDSGRAVQALAAVDEASRDIPRLRLVAIEPDDPGYAAAGAAAKRLAAENAV 68  
 QY 68 REAASAVEAAGKLVKIEETDIPRGAPVLVEASRDAMI CVSGVIGRYASSILGSTATTE 127  
 DB 69 RYATPAVDAADRPVKEVEIETQERVTSLIRASAAALAVCGAIGVNHFRERVGSTAA 128  
 QY 128 LAEYACFVAVMRSKVDPADINMIVVMTDAPDNEAVLEAAREAKLRQAPILALGCR 187  
 DB 129 LALSAQCCVAVIYVRPRVPIGRDAWIVVEADGSSDIGTLGAVVAEARLRDSPVAVVTGR 188  
 QY 188 PEELREIPDG-----GEFRRVODMHHNRPVAVYPIITHTGIGARPLADHDERVOLAVIGGSE 244  
 DB 189 QSGVGDTDDVRAFLDRWLARQPRYRVDVVOASAAVHGBELLDYLAGLRSVHMVVLLASD 248

QY 245 AGQLARLVGSGHVPFRHAECSVLV 270  
 DB 249 QEHVEQLVGAAGNAVLQEGACTLLV 274

RESULT 5  
 Q7TX68\_MYCBO  
 ID Q7TX68\_MYCBO PRELIMINARY; PRT; 268 AA.  
 AC Q7TX68;  
 DT 01-OCT-2003 (TReMBLrel. 25, Created)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)  
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
 DE Hypothetical protein MD3158c.  
 GN OrderedLocustNames=MD3158c;  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;  
 OC Mycobacterium tuberculosis complex.  
 NC NCB1\_TaxID=1765;  
 RX NUCLEOTIDE SEQUENCE.  
 RP STRAIN=AF2122/97;  
 RC MEDLINE=22709107; PubMed=12786972; DOI=10.1073/pnas.1130426100;  
 RA Garnier T., Bigliameier K., Camus J.-C., Medina N., Mansoor H.,  
 RA Pryor M., Dutfoy S., Gordin S., Lacroix C., Monsemp C., Simon S.,  
 RA Harris B., Alkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
 RT "The complete genome sequence of Mycobacterium bovis."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
 DR EMBL; BX248345; CAD95250.1; --; Genomic\_DNA.  
 DR GO; GO:0006950; P:response to stress; IEA.  
 DR InterPro; IPR006015; Unp.  
 DR InterPro; IPR006016; UnpA.  
 DR Pfam; PF00582; Unp; 1.  
 DR PRINTS; PR01438; UNYRS1STRESS.  
 KW Complete proteome; Hypothetical protein.  
 SO SEQUENCE 268 AA; 28008 MW; E82CB4F9C9D602AE CRC64;

Query Match 25.6%; Score 357.5; DB 2; Length 268;  
 Best Local Similarity 36.5%; Pred. No. 2.4e-19;  
 Matches 100; Conservative 35; Mismatches 110; Indels 29; Gaps 6;

QY 10 KTIIVGIDGSHAAITAAWGVDEAISRVAFLVSVIKPTHPSPDDYDRLDAHAESLRE 69  
 DB 8 RAVVVGIDGSRRAATHALMAVDEAVNRDIPRLRLVVIDPSQLSAAGCGGSAARALHD 67  
 QY 70 AQSAAVEAAGKLVKIEETDIPRGAPVLVEASRDAMI CVSGVIGRYASSILGSTATTE 129  
 DB 68 ASRRVEATGQPVKLETVLCGRPLTKMQSRSAAMLCVSGVGLD-HVGRGRGSVAATTLA 126  
 QY 130 EKAAICPVAVMRSKVDPDPA--SDINMIVVMTDAPDNEAVLEAAREAKLRQAPILALGCR 187  
 DB 127 GSAICPVAVINPSPAERATTSQGSAAVAEV-----DNEVLRHARPEERARLQVPLRAVA-- 180  
 QY 188 PEELREIPDG-----EFERRVODMHHNRPVAVYPIITHTGIGARPLADHDERVOL 237  
 DB 181 -VHAETFPDVEGQSRILAHYHLSRLAHMTRLYBEVAVDBAIAAGSACRHLAANAAXGQL 239  
 QY 238 AVTGGGAGQLARLVGSGHVPFRHAECSVLV 271  
 DB 240 FVADSHSAHELCGAYQPG-----CAVLTVR 264

RESULT 6  
 P95192\_MYCTU  
 ID P95192\_MYCTU PRELIMINARY; PRT; 268 AA.  
 AC P95192; Q7D624;  
 DT 01-MAY-1997 (TReMBLrel. 03, Created)  
 DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)  
 DT 13-SEP-2005 (TReMBLrel. 31, Last annotation update)  
 DE Hypothetical protein (Universal stress protein family).  
 GN OrderedLocustNames=MT3220, Rv3134c;  
 OS Mycobacterium tuberculosis.

```

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium tuberculosis complex.
OX NCBI_TaxID=1773;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230; DOI=10.1038/31159;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C.M.,
RA Harris D.E., Gordon S.V., Eigmeier K., Gae S., Barry C.E. III,
RA Tekala F., Badcock K., Basham D., Brown D., Chillingworth T.,
RA Connor R., Davies R.M., Devlin K., Fellwell T., Gentles S., Hamlin N.,
RA Holroyd S., Hornby T., Jagels K., Krogh A., McLean J., Moule S.,
RA Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,
RA Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
DOI=10.1128/JB.184.19.5479-5490.2002;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.D., DeBoy K.T., Dodson R.J., Gwinn M.L., Haft D.H.,
RA Hickey E.K., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D.,
RA Salzberg S.L., Delcher A.L., Uterback T.R., Weisman J.F., Khouri H.M.,
RA Gill J., Mikula A., Blahni W., Jacobs W.R. Jr., Venter J.C.,
RA Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
DR EMBL; BX842582; CAB06280.1; -: Genomic DNA.
DR PIR; G70645; G70645.
DR TIGR; MT3220; -.
DR Tuberculetic; RV3134C; -.
DR GO; GO:0006950; P:response to stress; IEA.
DR InterPro; IPR006015; Usp.
DR Pfam; PF00582; UspA.
DR PRINTS; PR01438; UNVSLSSTRESS.
DR Complete proteome; Hypothetical protein.
RW KMW
SQ SEQUENCE 268 AA; 28008 MW; EE2CB4F9C9D602AE CRC64;

Query Match 25.6%; Score 357.5; DB 2; Length 268;
Best Local Similarity 36.5%; Pred. No. 2,4e-19;
Matches 100; Conservative 35; Mismatches 110; Indels 29; Gaps 6;

QY 10 KTIIVGIDGSHAHTAALMGVDEAISRVAVPLRLVSVIKPTHSPDDYDRDLAAHRSURE 69
DB 8 RAVVVGIDGSRRAATTAALMAVDEAVNRDIPRLRVVVIDPSQSAAGEGGSAARALHD 67
QY 70 AQSVAEAGKLVKIEFDIPRGPAGVULVEASRDAMICVSGVIGRYASSILGSTATELA 129
DB 68 ASRKEVEAGQVQKIEVELCRPLTKMQESRSAAMLCVGSGVLD-HVRRGRGSVAATLA 126
QY 130 EKACPVAVMSKVDOP--SDIMNIVVMTDAPDNEAVLEVAAREALRQAPILALGCR 187
DB 127 GSALCPVAVVHPSPREPATTISQVSAVVAE---DNGVVLKHAPEEARLKGVPRLA-- 180
QY 188 PEELREIPDG-----EFERRVODMHHRHDPVRVYPTTHTGIRFLADHDERVOL 237
DB 181 -VHAAEPTDVEQSRRLAHVLSRLAHVTRLYPEVRDRAIAGSACRHLAANAKPGQL 239
QY 238 AVTGGGAGQTLARLVGSPGHVFRHAESVLYVR 271
DB 240 FVADSHSHEILGAYQPG-----CAVLTVR 264

```

```

ID Q73V37_MYCPA PRELIMINARY; PRT; 252 AA.
AC Q73V37;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=MAP3179c;
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OX NCBI_TaxID=1770;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K10;
RA Li L., Bannantine J., Zhang Q., Amornsri A., Alt D., Kapur V.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017238; AAS05727.1; -: Genomic DNA.
DR GO; GO:0006950; P:response to stress; IEA.
DR InterPro; IPR006016; UspA.
DR Pfam; PF00582; Usp; 1.
RW KMW
SQ SEQUENCE 252 AA; 26447 MW; FEDA63EB3EA80B1 CRC64;

Query Match 24.5%; Score 342.5; DB 2; Length 252;
Best Local Similarity 36.2%; Pred. No. 3,2e-18;
Matches 98; Conservative 41; Mismatches 89; Indels 43; Gaps 10;

QY 12 IIVGIDGSHAHTAALMGVDEAISRVAVPLRLVSVIKPTHSPDDY--DRDLAAHRSURE 69
DB 10 IIVGVDSRRAMAAVAVAIDAVGRDIPRLRVVVID-HGAPGSGHPTRLAAARALAD 68
QY 70 AQSVAEAGKLVKIEFDIPRGPAGVULVEASRDAMICVSGVIGRYASSILGSTATELA 129
DB 69 AHRVADFAQPVKVEITILWGNTAFKLLEOSRSAAVMLCVGQIGL-NHACGSPALATSLV 127
QY 130 EKACPVAVMSKVDOP--ASDIMNIVVMTDAPDNEAVLEVAAREALRQAPILALGCR 188
DB 128 RSALCPVAVVQOAPSLPAAARVSGVAEV---DNGVTLHAPEEARLKGVCAGVNP 183
QY 189 EELREIPDGSEFERRVODMHHRHDPVRVYPTTHTGIRFL-ADHDERVOLAVTGGGAGQ 247
DB 184 ARV-----ELERRLAMKMLRPVQAESAVLVGSGVGHRLADH-----RAGR 225
QY 248 L-----ARLVGSPGHVFRHAESVLYVR 271
DB 226 LVTVDAYRABALCHAGH-----SVLAVR 248

```

RESULT 7  
Q73V37\_MYCPA

DR EMBL, BX248341; CAD96904.1; -; Genomic\_DNA.  
 DR GO; GO:0006950; P:response to stress; IEA.  
 DR InterPro; IPR006015; Unsp.  
 DR InterPro; IPR006016; Unsp.  
 DR Pfam; PF00582; Unsp; 2.  
 DR PRINTS; PR01438; UNRS1STRESS.  
 DR Complete proteome; Hypothetical protein.  
 KW SEQUENCE 294 AA; 31456 MW; F0627782110F7142 CRC64;  
 SQ  
 Query Match 21.4%; Score 298.5; DB 2; Length 294;  
 Best Local Similarity 26.8%; Pred. No. 9.3e-15;  
 Matches 83; Conservative 53; Mismatches 119; Indels 33; Gaps 8;  
 QY 12 IIVGIDGSHAATTAALMGVDEAISRVAFLRLVSVIKPT-----HSPDDYDRDLAH 62  
 DB 10 ILVGVDSAGQNAVAAPAAAREAVNRQLPTLHIVAVVVGMPVQGLYANTEMQKD--N 67  
 QY 63 AERSLREAGSAVE---AAGKLVKIETDIPRGAPGVLEASRDAMTCVSGVIGRYASS 119  
 DB 68 AQQVIEQAREALTNSLSESKRPQVHTLVSNNVPTLLIDASQQLMVMVSGQMGALGRL 127  
 QY 120 ILGSTATELAKACFPVAVNRS-KVDOPASDINMIVMTDAPDNEAVLEVAAREAKLRQ 178  
 DB 128 LLGSIISTALLHARCPVAIIHSGNGATPDSAP-VLVGIDGSPASEATATAFDEARRR 186  
 QY 179 A-----PIALGGRPEELREIPDGEFERVODMHRHPDVRVYPITHTGI 224  
 DB 187 VDLVALHAWTDLGMPFLVGMWREREREKAE--VLAERLAGMQEQYPRVVRHRSLVCDKP 244  
 QY 225 ARFLADHDERVOLAVTIGG-GEAGQLARLVGSPGHVFRHAECVLYVR 271  
 DB 245 ARWLLEHSEQAQLVVGSHGRGFGSGMLDSVSSAVHVSRIPIVYR 292  
 RESULT 9  
 ID 053472 MYCTU PRELIMINARY; PRT; 294 AA.  
 AC 053472; Q7D7L7;  
 DT 01-JUN-1998 (TREMblrel. 06, Created)  
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
 DT 13-SEP-2005 (TREMblrel. 31, Last annotation update)  
 DE Hypothetical protein (Universal stress protein family).  
 GN OrderedlocusNames=MT2085, RV2026c;  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;  
 OC Mycobacterium tuberculosis complex.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230; DOI=10.1038/31159;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C.M.,  
 RA Harris D.E., Gordon S.V., Eigmeier K., Gae S., Barry C.E. III,  
 RA Tekala F., Badcock K., Basham D., Brown D., Chillingworth T.,  
 RA Connor R., Davies R.M., Devlin K., Fellwell T., Gentles S., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jagers K., Krogh A., McLean J., Moule S.,  
 RA Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,  
 RA Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sultson J.E., Taylor K., Whitehead S., Barrall B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RX MEDLINE=2206494; PubMed=12218036;  
 RA DOI=10.1128/DB.184.19.5479-5490.2002;  
 RA Flischmann R.D., Alland D., Bisen J.A., Carpenter L., White O.,  
 RA Peterson J.D., Debby R.T., Dodson R.J., Gwinn M.L., Haft D.H.,  
 RA Hickey E.K., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D.,  
 RA Salzberg S.L., Delcher A., Uterback T.R., Weisman J.F., Kouri H.M.,  
 RA Gill J., Mikhia A., Bishai W., Jacobs W.R. Jr., Venter J.C.,

RA Fraser C.M.;  
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL J. Bacteriol. 184:5479-5490 (2002).  
 DR EMBL, BX942578; CAA17240.1; -; Genomic\_DNA.  
 DR EMBL, AE000516; AKA46364.1; -; Genomic\_DNA.  
 DR PIR, A70942; A70942.  
 DR TIGR, MT2085; -;  
 DR Tuberculast; RV2026c; -;  
 DR GO; GO:0006950; P:response to stress; IEA.  
 DR InterPro; IPR006015; Unsp.  
 DR InterPro; IPR006016; Unsp.  
 DR Pfam; PF00582; Unsp; 2.  
 DR PRINTS; PR01438; UNRS1STRESS.  
 DR Complete proteome; Hypothetical protein.  
 KW SEQUENCE 294 AA; 31456 MW; F0627782110F7142 CRC64;  
 SQ  
 Query Match 21.4%; Score 298.5; DB 2; Length 294;  
 Best Local Similarity 26.8%; Pred. No. 9.3e-15;  
 Matches 83; Conservative 53; Mismatches 119; Indels 33; Gaps 8;  
 QY 12 IIVGIDGSHAATTAALMGVDEAISRVAFLRLVSVIKPT-----HSPDDYDRDLAH 62  
 DB 10 ILVGVDSAGQNAVAAPAAAREAVNRQLPTLHIVAVVVGMPVQGLYANTEMQKD--N 67  
 QY 63 AERSLREAGSAVE---AAGKLVKIETDIPRGAPGVLEASRDAMTCVSGVIGRYASS 119  
 DB 68 AQQVIEQAREALTNSLSESKRPQVHTLVSNNVPTLLIDASQQLMVMVSGQMGALGRL 127  
 QY 120 ILGSTATELAKACFPVAVNRS-KVDOPASDINMIVMTDAPDNEAVLEVAAREAKLRQ 178  
 DB 128 LLGSIISTALLHARCPVAIIHSGNGATPDSAP-VLVGIDGSPASEATATAFDEARRR 186  
 QY 179 A-----PIALGGRPEELREIPDGEFERVODMHRHPDVRVYPITHTGI 224  
 DB 187 VDLVALHAWTDLGMPFLVGMWREREREKAE--VLAERLAGMQEQYPRVVRHRSLVCDKP 244  
 QY 225 ARFLADHDERVOLAVTIGG-GEAGQLARLVGSPGHVFRHAECVLYVR 271  
 DB 245 ARWLLEHSEQAQLVVGSHGRGFGSGMLDSVSSAVHVSRIPIVYR 292  
 RESULT 10  
 ID Y2005 MYCTU STANDARD; PRT; 295 AA.  
 AC P64921; Q10851;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 25-OCT-2004 (Rel. 45, Last sequence update)  
 DT 10-MAY-2005 (Rel. 47, Last annotation update)  
 DE Hypothetical protein RV2005c/MT2061.  
 GN OrderedlocusNames=RV2005c, MT2061; ORFNames=MTCY39.12;  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;  
 OC Mycobacterium tuberculosis complex.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230; DOI=10.1038/31159;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C.M.,  
 RA Harris D.E., Gordon S.V., Eigmeier K., Gae S., Barry C.E. III,  
 RA Tekala F., Badcock K., Basham D., Brown D., Chillingworth T.,  
 RA Connor R., Davies R.M., Devlin K., Fellwell T., Gentles S., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jagers K., Krogh A., McLean J., Moule S.,  
 RA Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,  
 RA Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sultson J.E., Taylor K., Whitehead S., Barrall B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STAIN=CDC 1551 / Oshkosh;  
RX MEDLINE=22206694; PubMed=12218036;  
RX DOI=10.1128/JB.184.19.5479-5490.2002;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter J., White O.,  
RA Peterson J.D., DeBoy R.T., Dodson R.J., Gwinn M.L., Haft D.H.,  
RA Hickey E.K., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D.,  
RA Salzberg S.L., Delcher A.L., Utterback T.R., Feldman J.E., Khouri H.M.,  
RA Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,  
RA Fraser C.M.;  
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains".  
RL J. Bacteriol. 184:5479-5490(2002).  
CC -1 SIMILARITY: To B.subtilis yxiE.  
-----  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use as long as its content is in no way modified and this statement is not  
CC removed.  
-----  
CC EMBL: BX842578; CAA98383.1; -; Genomic DNA.  
CC EMBL: AE000516; AKK46338.1; -; Genomic\_DNA.  
CC PIR: C70759; C70759.  
DR TIGR: MT2061; -;  
DR Tuberculist; RV2005C; -;  
DR InterPro; IPR006015; Usp.  
DR InterPro; IPR006016; UspA.  
DR Pfam; PF00582; Usp; 2.  
DR PRINTS; PR01438; UNWRSLSTRESS.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 295 AA; 30985 MW; 827D60452824BE33 CRC64;

Query Match	20.5%;	Score 286;	DB 1;	Length 295;
Best Local Similarity	29.6%;	Pred. No. 8.6e-14;		
Matches	85;	Conservative 48;	Mismatches 126;	Indels 28;
			Gaps	8;

```

QY 12 IIVGDSGHAITLAWGDEALISAAPLRLTVSK-----PTHSPDDY-----DSDL 60
Db 10 VVVGDSGLSEBPAAKCMGATDPAAMRNIFUTVVHVNADVAIYMPMPYETGMCWOEDGCR 69
QY 61 AHAERLREAGSAVEBAKLVYKTIETDIPRGPAPVLVBSRDAEMI CVGSGVIGRYASSI 120
Db 70 QIVANAIVKLAEAVGADRKL-SVKSELYSTFVPMTVEIISNEAEVNVLGSSGRCALAGL 128
QY 121 LGSTALEAKRAHCVAIYMRSKVQDPADINM--IYVMTDAPDNEAVLEAARBAKLRQ 178
Db 129 LGSVSSSLVRRRGCCVAAVISHD-DAVIPDPQHAPALVIGDGSVSELAITAVAFDSASRG 187
QY 179 APIILAGRPE-ELREIIPDEGE-----RRVDMHRRHBDVVRPIYTHHTGJAR 226
Db 188 VELLIVAHMSDVEVELFGDPSAVQGEAELESLARLGMQGERYDVPVSRVVCDDRPAR 247
QY 227 FLADDERVOLAVTGG-GEAGOLARLVBPBGHPVRNHAECSLVYVR 272
Db 248 KLVOKSASQOLVVVSGHGRGILTGMLLGSVSAVLAHARVPYIYARQ 294

```

```

RESULT 11
Y2028_MYCBO
ID Y2028_MYCBO STANDARD, PRT; 295 AA.
AC P64922; Q10851;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Hypothetical protein M2028c.
GN OrderedLocustNames=M2028c;
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium tuberculosis complex.
CX NCBI_TaxID=1765;
NP [1]
NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

```

RC STBAINE=AF2102197; PubMed=12768972; DOI=10.1073/pnas.1130426100;  
RX MEDLINE=227030107; PubMed=12768972; DOI=10.1073/pnas.1130426100;  
RA Garnier L., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,  
RA Puyot M., Dutfoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,  
RA Harris B., Ackin R., Doggett U., Mayes R., Keating L., Wheeler P.R.,  
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,  
RT "The complete genome sequence of *Mycobacterium bovis*,"  
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
CC -1- SIMILARITY: To B.subtilis yx1E.  
CC -----  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use as long as its content is in no way modified and this statement is not  
CC removed.  
CC -----  
DR EMBL, BX248341, CAD96881.1; -, Genomic DNA.  
DR InterPro, IPR006015; Usp.  
DR InterPro, IPR006016; Uspa.  
DR Pfam; PF00582; Usp; 2.  
DR PRINTS; PR01438; UNIVRS.LSTRESS.  
KM Complete proteome; Hypothetical protein.  
SQ SEQUENCE 295 AA; 30985 MW; 8275D0452E24B33 CRC64;

Query Match	20.5%;	Score 286;	DB 1;	length 295;
Best Local Similarity	29.6%;	Pred. No. 8.6e-14;		
Matches	85;	Conservative	48;	Mismatches 126;
			Indels	28;
			Gaps	8;

[illegible]

RESULT 12	Q6FLR9_COREF	Q6FLR9_COREF PRELIMINARY;	PRF;	318 AA.
AC	Q6FLR9;			
DT	01-MAR-2003 (TrEMBLrel. 23, Created)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)			
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)			
DE	Hypothetical protein.			
CN	OrderedLocustNames=CE2791,			
OS	Corynebacterium efficiens.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.			
OX	NCBI_TaxID=152794;			
RN	[1]			
RP	STRUCTURE SEQUENCE.			
RC	STRAIN=Y8-314 / AT 12310 / DSM 44549 / JCM 11189;			
RX	MEDLINE=22723752; PubMed=12840036; DOI=10.1101/gr.1285603;			
RA	Nishio Y., Nakamura Y., Kawarayashi Y., Usuda Y., Kimura E.,			
RA	Sugimoto S., Matsui K., Yamagishi A., Kikuchi H., Ikeo K.,			
RA	Gotohori T.;			
RT	"Comparative complete genome sequence analysis of the amino acid			
RT	replacements responsible for the thermostability of Corynebacterium			
RT	efficiens.";			
RL	Genome Res. 13:1572-1579 (2003)			







Db 243 RPARQLIQSSSAQLTVVGSHGRLAGTLLGSVSNNAVHVSVPYIVAR 292

RESULT 15

ID Q7TY75\_MYCBO PRELIMINARY; PRT; 297 AA.  
AC Q7TY75;  
DT 01-OCT-2003 (TREMBLrel. 25, Created)  
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Hypothetical protein TB31.7.  
GN Name=TB31.7; OrderedLocusName=Mb2656;  
OS Mycobacterium bovis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;  
OC Mycobacterium tuberculosis complex.  
CX NCBI\_TaxID=1765;  
NM [1]  
NP NUCLEOTIDE SEQUENCE.  
RC STRAIN=AF2122/97;  
RX MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;  
RA Garnier T., Sigtleier K., Camus J.-C., Medina N., Mansoor H.,  
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempé C., Simon S.,  
RA Hatris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
RT "The complete genome sequence of Mycobacterium bovis."  
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
DR EMBL: BX248343; CAD94841.1; -: Genomic DNA.  
DR GO: GO:0006950; P:response to stress; IEA.  
DR InterPro: IPR006015; USP.  
DR InterPro: IPR006016; USPA.  
DR Pfam: PF00582; USP; 2.  
DR PRINTS: PR01438; UNRSLSSTRESS.  
KW Complete proteome.  
SQ SEQUENCE 297 AA; 31652 MW; A590F7058D1E8695 CRC64;

Query Match 19.1%; Score 267; DB 2; Length 297;  
Best Local Similarity 29.3%; Pred. No. 2.5e-12;  
Matches 84; Conservative 42; Mismatches 131; Indels 30; Gaps 8;

QY 12 IIVGDSHAITATLWGVDEAISRVLRYSVIKP-----THRSP-----DDYDR 58  
DB 10 IIVGIDDSPPAAQVAVRWAPARDAELRKIPLTIVHAVSPVATWLEVPPLPGVLRWQDDHR 69  
QY 59 DLAAERSLREASQSAVEAAGKLVKLTETDIPRGAPVIVEASRDALMICVSGVIGRYAS 118  
DB 70 HL--IDDLKLVVEQASLRAPG-PTVHSEIVPAAAVPTLVMSKDAVLMVVGCLSGGRWPG 126  
QY 119 SILGSTATELAEKACPVAVNRKSKVD-QPASDINMIVVMTDAPDNEAVLEVAAREAKLR 177  
DB 127 RLTSVSSSGLRHAHCPVVIIDEDSVMPHPOQAPVLGVGDSSASSELATATAIFDEASRR 186  
QY 178 QAPILALGRPE-ELREIP-----DGEFERVQDWHHRHDPVRYVPIITHTGIA 225  
DB 187 NVDLVVALHAWSDVDVSEWPGIDWPATQSMAEQLAERLAGWQERYPNVAITRVVVRDQPA 246  
QY 226 RFLADHDERVOLAVIGG-GEAGQLARLVGPSGHPVFRHAECSTLVVR 271  
DB 247 RQLVORSEEAQLVVVGSRGGRGYAGMLVGSVGETVTAQLARTPTVIYAR 293

Search completed: March 23, 2006, 05:24:26  
Job time : 70.5674 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

## OM protein - protein search, using BW model

Run on: March 23, 2006, 05:24:53 ; Search time 18.5513 Seconds  
(without alignments)  
1212.194 Million cell updates/sec

Title: US-10-617-038-27  
Perfect score: 1397  
Sequence: 1 MSGRGPVTKTIIVIGIDSH.....GPSGHPVPRHAECSVLVRR 272

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/6\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/H\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/PCitus\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/RS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	266	19.0	301	2	US-09-605-703B-1870
2	112	8.0	189	2	US-09-134-001C-3795
3	105.5	7.6	613	2	US-09-758-759-151
4	105	7.5	151	2	US-09-107-532A-5606
5	103.5	7.4	430	2	US-09-902-540-12843
6	100.5	7.2	1028	2	US-09-328-352-5749
7	98.5	7.1	542	2	US-09-252-991A-37590
8	98.5	7.1	735	2	US-09-252-991A-30569
9	98	7.0	657	2	US-09-489-039A-10221
10	97.5	7.0	715	2	US-09-902-540-12939
11	97	6.9	148	2	US-09-107-433-3293
12	97	6.9	150	2	US-09-583-110-3588
13	97	6.9	161	2	US-09-107-532A-3842
14	97	6.9	420	2	US-09-252-991A-10438
15	94	6.7	157	2	US-09-328-352-7302
16	94	6.7	800	2	US-09-252-991A-11975
17	93.5	6.7	531	2	US-09-252-991A-17650
18	93.5	6.7	693	2	US-09-252-991A-19147
19	93	6.7	485	2	US-09-489-039A-10554
20	93	6.7	565	2	US-09-252-991A-23494
21	92.5	6.6	366	2	US-09-252-991A-12542
22	92	6.6	367	2	US-09-902-540-14540
23	92	6.6	844	2	US-09-252-991A-27184
24	91.5	6.5	331	2	US-09-252-991A-29383
25	91.5	6.5	452	2	US-09-252-991A-18948
26	90.5	6.5	431	2	US-09-252-991A-20267
27	90	6.4	467	2	US-09-252-991A-23045

28	89.5	6.4	478	2	US-09-252-991A-31866	Sequence 31866, A
29	89.5	6.4	484	2	US-09-252-991A-26896	Sequence 26896, A
30	89.5	6.4	585	2	US-09-252-991A-30045	Sequence 30045, A
31	89.5	6.4	615	2	US-09-949-016-11320	Sequence 11320, A
32	89.5	6.4	684	2	US-09-252-991A-27004	Sequence 27004, A
33	89.5	6.4	1388	2	US-09-252-991A-20237	Sequence 20237, A
34	89.5	6.4	1953	2	US-09-917-254-92	Sequence 92, Appl
35	89	6.4	512	2	US-09-902-540-15789	Sequence 15789, A
36	88.5	6.3	1421	2	US-09-335-409-2	Sequence 2, Appl1
37	88.5	6.3	1421	2	US-09-568-102-2	Sequence 2, Appl1
38	88.5	6.3	1421	2	US-09-567-969-2	Sequence 2, Appl1
39	88.5	6.3	1421	2	US-09-568-480-2	Sequence 2, Appl1
40	88.5	6.3	1421	2	US-09-568-486-2	Sequence 2, Appl1
41	88.5	6.3	1421	2	US-09-568-472-2	Sequence 2, Appl1
42	88.5	6.3	1421	2	US-09-567-899-2	Sequence 2, Appl1
43	88.5	6.3	1421	2	US-10-014-717-2	Sequence 2, Appl1
44	88	6.3	403	2	US-09-489-039A-11881	Sequence 11881, A
45	87.5	6.3	547	2	US-09-252-991A-18576	Sequence 18576, A

## ALIGNMENTS

```
RESULT 1
US-09-605-703B-1870
; Sequence 1870, Application US/09605703B
; Patent No. 6962989
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Krieger, Burkhard
; APPLICANT: Schröder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauser, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL
; TITLE OR INVENTION: PROTEINS
; FILE REFERENCE: BG1-129CP
; CURRENT APPLICATION NUMBER: US/09/605, 703B
; PRIOR FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: 60/142,764
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: 60/152,318
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 2934
; SEQ ID NO 1870
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-605-703B-1870
Query Match 19.0%; Score 266; DB 2; Length 301;
Best Local Similarity 28.8%; Pred. No. 7.7e-21;
Matches 85; Conservative 45; Mismatches 11; Indels 34; Gaps 8;
QY 8 TMKTIIVIGIDSHAAITPAALMGVDEAISRAVPLRLVSIVIKETPSS-----PDDYDR 58
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 3 TEDIIVAVVGDSDASKAVRMAANTANKRGIPLELAS--SYTWPOFLYAEGMVPPQLFD 60
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 59 DL-AHARSRLREKQSAVEAAGKVKITETDIPREGAGVLYEASDAMICVSGIGIRYA 117
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 61 DLQEALEKINEARDIAHEVAPELKIIGHTIAEGSPIDMLLEMSPDATWIVGSGLSGLS 120
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 118 SSIIIGSTATLEAGVACPCPAVMR--SKYDOPASDINNIYVMTDPAPNEAVLEVAAREAK 175
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 121 GMYGSGVAGVAVSHAKCPVVVVRKDSAVNED--SKYGVVVGVDGSEVSGQATEVAFPAE 179
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 176 LR-----QAPILAGRPPELREIPDGEFE--RRVODMHRHPDVRYVP 217
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 180 ARGAEIVAVHTWMDMVQASIIAGAAQOOOWDEVERQOTMLTERLAPLVEKYPSTVTKK 239
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 218 ITTHTTGARPLADDERVQGLAVTGG--GEAGQLARLVPSGHPVPRHAECSVLVVR 271
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 240 ITTRDRPVRALAEKSENAQLLVGSGHGGPKGMLLASTSRALQSPAPCPMNVVR 294
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
```



```

Db      13  QNTIMVAVDGRQSIKAFALBDLAKONEAHLPIVSIINKVELTSHSAFXYIADKQRT  72
        70  AQGSVRLAAGLVK-----IENDIPGRGAPLVLEASDAEMICGSGIGRYASII  120
        :::::|||||:::
Db      73  EVALLKKKINDAKYEGIAETHAIVETDPPNLTIANV-IFGEQENIDILVIGATGKCAIQOL  131
        :::::|||||:::
Qy      121  LGSATELEAKAHCPVAVNR  140
        ::|||::|::|::|
Db      132  VGSFASIVVSHAPCSVLVVK  151
        ::|||::|::|::|

```

RESULT 5  
US-09-902-540-12843

```

GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10 (15849) B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 12843
LENGTH: 430
TYPE: PRT
ORGANISM: Myxococcus xanthus
IS-09-902-540-12843

```

Query Match	7.4%	Score	103.5	DB	2	Length	430
Best Local Similarity	24.2%	Pred. No.	0.014				
Matches	78	Conservative	25	Mismatches	118	Indels	101
						Gaps	14

Qy	1	TIIVGDGSHAAITTAALMVIDEASBAVPLRVSYIKTTHSPDDYURD-----LAAAR	65
Db	2	SIYCAINFDMAARRASTLAAELAKKAKGISTLMVHYLN-----DSVAFKALLSGREA	55
Qy	66	SIREASAAVEAMGKLVKIEITDI.PRG-PA-----GPVLVEASDAEMI CVGVSIG	114
Db	56	VLEDETKRLEQLG--KVPEPVLLTGEPVAVLEGFCEBOGASLVVASRAADESPGGE---	110
Qy	115	RYASSTLGSATETALAKACPVAVNR-----	140
Db	111	-----GGVDDMAOSLTVPLLVNRDPRVLEAMVGRGERSLKVFGVDRSLSPFEAADWV	163
Qy	141	-----SKVDORASDINMTIVRMTDAPDNEAVLEUYAAAEAKLRQAPILALGGRPEELRET	194
Db	164	LTLSSKGRVDVUGRVLM-----PEEE-----ARLGLRQ--LAFDALPELQHV	207
Qy	195	RODEFERVRQDMHNRADVRVVRITTHNTSLARFLADH-----DERVQLAVIGGEGAGOL	248
Db	208	LEBECSBELLARLAOGSTPVRA--RVEMKVGK-IDHLLVDLADQENHDLVVGTHNRAL	263
Qy	249	ARLVGSGHVPFHAECSVLVV	270
Db	264	GRLMVSRR--ALRIAMSVVVCV	284

RESULT 6  
US-09-328-352-5749  
; Sequence 5749, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04

```
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5749
; LENGTH: 1028
; TYPE: PRF
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5749
```

Query Match	7.2%;	Score 100.5;	DB 2;	Length 1028;
Best Local Similarity	25.7%;	Pred. No. 0.12;		
Matches	54;	Conservative	33;	Mismatches 88;
				Indels 35;
				Gaps 12

```

OY      22  ATRPAAAGVDEALSRAPVRLTSVVKPTHPSPDDYDRODLAAEESLRACQAVAACTLV  81
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB      749  AVSALMCMVNMEMR---YKMSPIKIRLTS--DYNRKV---EELALNGEDLIPTWKPS  800
OY      82  KIETD--IIP-GAGGVVLVEASRDMEI CVSGVIGRASSLSIGSTTELAEXA-----H  133
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB      801  DSAIGERAPPLTPSIVTVADEPDMT---MVGQKAEMET---TRLQKSRPAAGIH  855
OY      134  CPVAVNRSKVDQ-----PASDINMTVNRMTA PDNEAVLEAYABEAKLRQAPILALG3--  186
DB      853  LLALATRPSTVDTGLIKANIPTRALVAVNSKIOSRTILIDGAGEDLLGHGDMFLPQK  912A

```

```

Qy      187 -RPEELRE--IPDGEFFERRVQDMHNR-HPD 2122
      || | | | | | | | |
Db      913 IEPERVHGAFLSDDEVNRICDAMRERGEED 9422

```

```

RESULT 7
US-09-252-991A-27590
; Sequence 27590, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27590
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27590

```

Query Match	7.1%;	Score 98.5;	DB 2;	Length 542;
Best Local Similarity	19.4%;	Pred. No. 0.073;		
Matches 63;	Conservative 45;	Mismatches 117;	Indels 99;	Gaps 13;

QY	10	KTIVIGD---	SSHAIIVAMIGVBEALSKRAFLNL	---	VSIAKTHSPUDYRDL	60
Db		158	RAIVIGGKAGAMEVVERKEHQOGEI	IGLVAPYRYGAQCSR	IEVEASHVPPDAGERV	21.7
QY	61	AHAERSIRENOASVANAAGLVKVIET	---	DIPRGAG---	---	PVYVASHDAE 10.4
Db	218	A---RRVLETISGLANSDVIFIL	GGSGSALLSLP	AEIGIGLADKOAVNKALLKSGA	IGE	27.4
QY	105	MICV---G	SVGIGRYA---	---	SSILGSTATELAE---	KHCEVAV 13.8
Db	275	MNCVRGHL	SAIKGRILAKACMPASV	FYTAISVDPED	ETVIAASGSPVGDPTT	SADALAIL 33.4
QY	139	MRSKVDPQPSDINWIVYMTDA-	PDN---	---	EAVLEVAAREAKLRQAPIL	16.2
Db	335	RRVAIVDPQVADIMLADP	RESETKRDPDDCLARSHFQ	LIATPQNALDDAAV	KCRAPG	STL 39.4
QY	183	ALGGRPEELREI	PDGEFERRVQDMHRRHPDV	RVVPEITTTHTG	IAIFLADHDE-	RVQALAVI 24.0
Db	395	ILGDLDEGARREV---	---	AKVNAGIAQOIRNGQPLA	PCVIL	42.1

Qy 241 GCGAGQALRVPSGHPFRHAE 264  
|||  
Db 432 SGGTTVTRNGRGGRG---RNAE 451

## RESULT 8

US-09-252-991A-30569  
; Sequence 30569, Application US/09252991A  
; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 30569

; LENGTH: 735

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-30569

Query Match 7.1%; Score 98.5; DB 2; Length 735;

Best Local Similarity 24.5%; Pred. No. 0.12; Mismatches 78; Indels 83; Gaps 14;

Matches 63; Conservative 33; Mismatches 78; Indels 83; Gaps 14;

Qy 5 GEPTMTIIVIGIDGSHAAITAAALMGVDEAISRVAFLVSVIKPTHPSPDDYDRDLAAE 64

Db 297 GAPAQ-LVLAQGAELVIAAG--ALDVLQABEPSRLVGV-----AE 335

Qy 65 RSLAEAGS-AVEAAGKLVKLTETDIPRGAGPYLVASDAEMICY-GSVGIGRYASSILG 122

Db 336 QVLLAAVALAAVAAAGR---QAEAGSGHGDV---AADACQAVAGGLGVGSGQAVAD 387

Qy 123 STATELEAKACPVAVMSKYDOPASDINMIVMTDAPDNEAVLEVAAREAKL---RQ 178

Db 388 AVAADADG---VVGIGVAAQ-----AVVAAQAPDQVETIAQAAKQGVYVERQ 434

Qy 179 A-----PILALGGRPE-----ELREIPDGEFERRVQDMHNRHPRVYVITHTGI--A 225

Db 435 AGHHLSTGVGSHHGRFAGSVDRHNVADG-----GTGLQVGA 472

Qy 226 RFLADHDERVQAVIYG 242

Db 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Qy 473 AGLVAVDEQVDGAGYGG 489

Query Match 7.0%; Score 98; DB 2; Length 657;  
Best Local Similarity 22.5%; Pred. No. 0.11;  
Matches 71; Conservative 47; Mismatches 108; Indels 90; Gaps 14;

Qy 5 GEPTMTIIVIGIDGSHAAITAAALMGVDEAISRVAFLVSVIKPTHPSPDDYDRDLAAE 64

Db 291 GEPIMQV-----TSALIRHMAHCDNDNOSQAAALIGIRHLRLQVLAAGLIKSRRP 344

Qy 51 PSPDDYDRDLAAERSLR-----EAGSAVEAAGKLVKLTETDIPRGAGPYLV 98

Db 345 PAPRAPANAGADRELIRIGYGRFSGILIKARQSLTETAPASIGVNVLMSEFPAGQLH 404

Qy 99 ASRAEMICVSGV-----IGRYASSILGSTAE-LAEKACPVAVMSKYDQ----- 145

Db 405 ALACNE-IDFGTGEAPVPAQASNSSELMYVAMEPPRAVAVVQGESDIRQLSDLRGK 463

Qy 146 -----PASDINMIV-----VRMTDAPDNEAVL--EYAAERATLQAPITALG 185

Db 464 RIALKNSVNHMLLQLEDAAGLADYRVYTPPKYPLTASDYLAVDAMMMDDPLLSDA 523

Qy 186 GRPELEIPDGE-----FERRVODMHRHPRV--RVYPTHTGIRAFDLADHDERVQ 236

Db 524 EHTGELRVASGGRVNNHQFYLRSRDYLAQHDIMRRLTETHTG--QFIDSH----- 576

Qy 237 LAVTGGGAGQALRV 252

Db 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Qy 577 -----RGEAARLT 584

Db 645 ARHGD-----LVIVIGAGMGLEDKVFGLQREIRIIDAPGSLIV 695

## RESULT 11

US-09-107-433-3293  
Sequence 3293, Application US/09107433  
Patent No. 6800744

## GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE

## THERAPEUTICS FOR DIAGN

NUMBER OF SEQUENCES: 5206

## CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA

ZIP: 02354

COUNTRY: USA

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: <Unknown>

OPERATING SYSTEM: <Unknown>

SOFTWARE: <Unknown>

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,433

PRIOR FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085131

PRIOR FILING DATE: May 12, 1998

APPLICATION NUMBER: 60/051553

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-011

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 3293:

SEQUENCE CHARACTERISTICS:

LENGTH: 148 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Streptococcus pneumoniae

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...148

SEQUENCE DESCRIPTION: SEQ ID NO: 3293:

US-09-107-433-3293

Query Match 6.9%; Score 97; DB 2; Length 148;

Best Local Similarity 25.4%; Pred. No. 0.014;

Matches 36; Conservative 27; Mismatches 69; Indels 10; Gaps 4;

Db

Qy 10 KTIIVIGDGHAAITAAIMGVDAISRVPRLVSVIKP-THPSDDYDRDL-----AHA 63

Db 4 QNIVVAIDGSKEDLAFVKGVSHALRNDAKLTIAHVIDTRALQSVSTFDAYEVELQYDA 63

Qy 64 ERSIREAQSVAEAGKLVKLTETDIPRGAPVLYEASRDAB---MICVSGVIGRYVSSI 120

Db 64 ESLMKEYEKAKDAG-VADVHIVTEMGNPKTLARTIPDAEBVDLIVGATGNAFERLL 122

Qy 121 LGSTATLAEKRAHCPVAVMSK 142

Db 123 VGSSEYILRRAKVDLVVREQ 144

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

US-09-583-110-3588

Sequence 3588, Application US/09583110

Patent No. 6699703

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al.

TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus

FILE REFERENCE: PATH00-07A

CURRENT APPLICATION NUMBER: US/09/583,110

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/107,433

PRIOR FILING DATE: 1998-06-30

PRIOR APPLICATION NUMBER: US 60/085,131

PRIOR FILING DATE: 1998-05-12

PRIOR APPLICATION NUMBER: US 60/051,553

PRIOR FILING DATE: 1997-07-02

NUMBER OF SEQ ID NOS: 5322

SEQ ID NO 3588

LENGTH: 150

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-583-110-3588

Query Match 6.9%; Score 97; DB 2; Length 150;

Best Local Similarity 25.4%; Pred. No. 0.014;

Matches 36; Conservative 27; Mismatches 69; Indels 10; Gaps 4;

Db

Qy 10 KTIIVIGDGHAAITAAIMGVDAISRVPRLVSVIKP-THPSDDYDRDL-----AHA 63

Db 6 QNIVVAIDGSKEDLAFVKGVSHALRNDAKLTIAHVIDTRALQSVSTFDAYEVELQYDA 65

Qy 64 ERSIREAQSVAEAGKLVKLTETDIPRGAPVLYEASRDAB---MICVSGVIGRYVSSI 120

Db 66 ESLMKEYEKAKDAG-VADVHIVTEMGNPKTLARTIPDAEBVDLIVGATGNAFERLL 124

Qy 121 LGSTATLAEKRAHCPVAVMSK 142

Db 125 VGSSEYILRRAKVDLVVREQ 146

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

```
TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3842:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 161 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (B) LOCATION 1..161
; SEQUENCE DESCRIPTION: SEQ ID NO: 3842:
US-09-107-532A-3842
```

```
Query Match      6.9%; Score 97; DB 2; Length 161;
Best Local Similarity 25.2%; Pred. No. 0.016;
Matches 38; Conservative 30; Mismatches 71; Indels 12; Gaps 5;
```

```
Qy      10 KTIIVGIDGSHAAITLALMGVDEAISRVPRLVSVIKP--THPSDDYDRDLA-----A 63
Db      8 KKIIVAVDGSDEALAFKAVVAVAIKRNNGELVLAHVIDTTSFQVSSFDGMLAEQATBMA 67
Qy      64 ERSIREAQSVAEAGKLVKLETIDIPRGAPVL---VEASRDAMICVSGVIGRYASSI 120
Db      68 KQTLADVESNKKAG--LNNVTSVVEYSGPKQIIAREIPEDNQVDLIMLGAGLNAVERLF 126
Qy      121 LGSTATLEAKACPCVAVMSKVDQ--PASD 149
Db      127 IGSVSEYVIRNAACDVAVVTRDTLENOLPAKD 157

RESULT 14
US-09-252-991A-30428
; Sequence 30428, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30428
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30428
```

```
Query Match      6.9%; Score 97; DB 2; Length 420;
Best Local Similarity 25.4%; Pred. No. 0.072;
Matches 58; Conservative 24; Mismatches 74; Indels 72; Gaps 13;
```

```
Qy      51 PSPD-DYDR-DLAHRSLEAQSVAEAGKLVKLETIDIPRGA-----GPVLVEASR 101
Db      216 PAPERNLDRPELNAVAHL-DAQARQVFLG-----DSPGHAHGGLARRGTAAVA 267
Qy      102 DAEMICVSGVIGRYASSIIGSTATLEAKACPCVAVMSKVDQPSADIMIVKMTDAP 161
Db      268 HAVLVVGVVGMGR-AEQVLDRRV-----LGLLVGVAQQA-----DRR 306
Qy      162 DNEAVLEVARBEAKLRQAPILALGGRP-----EELREIPDGEFERRVQDMHHRPDV 213
Db      307 AGGLALEHGRK--NLHLVGFALGSDVPAAGLAEVVALQVPGGQFQPRGA----- 355
```

```
Qy      214 RVPYITHTGIARFLADHDERVOLAVYGGEGAGQLARLVPSGHPVR 261
Db      356 ---PID-----DGDORRAVAFACGGDGBELA--VGVAGHARLR 388
```

```
RESULT 15
US-09-328-352-7302
; Sequence 7302, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7302
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7302
```

```
Query Match      6.7%; Score 94; DB 2; Length 157;
Best Local Similarity 25.4%; Pred. No. 0.033;
Matches 36; Conservative 27; Mismatches 63; Indels 16; Gaps 5;
```

```
Qy      12 IIVGIDGS-----HAITLALMGVDEAISRVP--PLRLVSVIKPTHPSDDYDRDL 60
Db      16 ILVPVDSPTSLAVNHMAASLAKAFSSKTVLVVALTIIDPISVEIIDSITEIADYENKAR 75
Qy      61 AHAERSLEAQSVAEAGKLVKLETIDIPRGA--GPVLVEASRD--AEMICVSGVIGRYA 117
Db      76 ASIGSILDOAKEQFSOHG--ISVETKIVGGQTIHTEIIRKATLAKADLLVIGSHGKGRK 133
Qy      118 SSILGSTATLEAKACPCVAVM 139
Db      134 KPFLGSVTQALLGEIHVPVLV 155
```

```
Search completed: March 23, 2006, 05:30:16
Job time : 19.6513 secs
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 06:14:33 ; Search time 57.1999 Seconds  
(without alignments)  
1986.885 Million cell updates/sec

Title: US-10-617-038-27

Perfect score: 1397  
Sequence: 1 MSGSGEPTMTTIVGIDGSH.....GPSGHPVFRHAECSVLVRR 272

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_AA\_Main:  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBSCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBSCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBSCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBSCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBSCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBSCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1397	100.0	272	US-10-617-038-27	Sequence 27, Appl
2	416	29.8	279	US-10-617-038-21	Sequence 21, Appl
3	357.5	25.6	268	US-10-617-038-42	Sequence 42, Appl
4	267	19.1	297	US-10-138-473-34	Sequence 34, Appl
5	267	19.1	297	US-10-450-726-4	Sequence 4, Appl
6	266	19.0	301	US-09-738-626-6745	Sequence 6745, Ap
7	266	19.0	301	US-09-746-660A-6	Sequence 6, Appl
8	266	19.0	301	US-09-746-660A-20	Sequence 20, Appl
9	266	19.0	301	US-11-050-055-8	Sequence 8, Appl
10	266	19.0	301	US-11-050-098-314	Sequence 314, App
11	249	17.8	300	US-10-283-122A-53861	Sequence 53861, A
12	242.5	17.4	290	US-10-156-761-8814	Sequence 8814, Ap
13	210	15.0	295	US-10-156-761-8839	Sequence 8839, Ap
14	143.5	10.3	185	US-10-437-963-153439	Sequence 153439,
15	130.5	9.3	163	US-10-424-559-154598	Sequence 154598,
16	129	9.2	176	US-10-767-701-40880	Sequence 40880, A
17	128.5	9.2	171	US-10-425-115-283083	Sequence 283083,
18	128.5	9.2	173	US-10-425-114-61762	Sequence 61762, A
19	125	8.9	179	US-10-437-963-128807	Sequence 128807,
20	124	8.9	165	US-10-928-992-74	Sequence 74, Appl
21	116.5	8.3	169	US-10-437-963-133809	Sequence 133809,
22	116	8.3	165	US-10-767-701-31566	Sequence 31566, A
23	115	8.2	160	US-10-437-963-126295	Sequence 126295,
24	114.5	8.2	181	US-10-437-963-126675	Sequence 126675,
25	112	8.0	189	US-10-724-972A-6081	Sequence 6081, Ap
26	109.5	7.8	163	US-10-424-599-215538	Sequence 215538,
27	109	7.8	165	US-10-425-115-228222	Sequence 228222,

28	109	7.8	941	4	US-10-282-122A-62427	Sequence 62427, A
29	109	7.8	941	4	US-10-282-122A-64612	Sequence 64612, A
30	108	7.7	147	3	US-09-738-626-5013	Sequence 5013, Ap
31	108	7.7	183	4	US-10-437-963-108194	Sequence 108194,
32	108	7.7	245	4	US-10-425-114-62796	Sequence 62796, A
33	108	7.7	256	4	US-10-425-115-305753	Sequence 305753,
34	107.5	7.7	177	4	US-10-424-599-170466	Sequence 170466,
35	107.5	7.7	6751	5	US-10-760-493-25	Sequence 25, Appl
36	107	7.7	206	4	US-10-425-114-47029	Sequence 47029, A
37	105.5	7.6	613	3	US-09-758-759-151	Sequence 151, App
38	105.5	7.6	613	6	US-11-021-825-151	Sequence 151, App
39	105	7.5	165	4	US-10-437-963-148540	Sequence 148540,
40	104.5	7.5	147	4	US-10-282-122A-62210	Sequence 62210, A
41	104.5	7.5	147	4	US-10-282-122A-63851	Sequence 63851, A
42	104.5	7.5	194	4	US-10-156-761-8452	Sequence 8452, Ap
43	103	7.4	174	4	US-10-425-115-222011	Sequence 222011,
44	103	7.4	182	4	US-10-424-599-170465	Sequence 170465,
45	102	7.3	164	4	US-10-425-115-242323	Sequence 242323,

## ALIGNMENTS

RESULT 1  
US-10-617-038-27  
Sequence 27, Application US/10617038  
Publication No. US20040057963A1  
GENERAL INFORMATION:  
APPLICANT: Andersen, Peter  
APPLICANT: Rosenkrands, Ida  
APPLICANT: Strlyhn, Anette  
TITLE OF INVENTION: Therapeutic TB Vaccine  
FILE REFERENCE: SSI5AUSA  
CURRENT APPLICATION NUMBER: US/10/617, 038  
CURRENT FILING DATE: 2003-07-11  
PRIOR APPLICATION NUMBER: DK PA 2002 01098  
PRIOR FILING DATE: 2002-07-13  
PRIOR APPLICATION NUMBER: US 60/401, 725  
PRIOR FILING DATE: 2002-08-07  
NUMBER OF SEQ ID NOS: 187  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 27  
LENGTH: 272  
TYPE: PRT  
ORGANISM: Mycobacterium tuberculosis  
US-10-617-038-27

Query Match 100.0%; Score 1397; DB 4; Length 272;  
Best Local Similarity 100.0%; Pred. No. 1.9e-125;  
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSGSGEPTMTTIVGIDGSHATPAALMGVDEAISRAVPLVSVIKPTHSPDDYRD	60
DB	1	MSGSGEPTMTTIVGIDGSHATPAALMGVDEAISRAVPLVSVIKPTHSPDDYRD	60
QY	61	AHAERSREAOAAYEAAGKVIETDIPRGAPVTLVEASRDAMICVSGVIGRYASSI	120
DB	61	AHAERSREAOAAYEAAGKVIETDIPRGAPVTLVEASRDAMICVSGVIGRYASSI	120
QY	121	LGSATLAEKACPCVAVMRSKVDOPASDINWIVYRTDAPDNEAVTEYAREKKAQAP	180
DB	121	LGSATLAEKACPCVAVMRSKVDOPASDINWIVYRTDAPDNEAVTEYAREKKAQAP	180
QY	181	ILALGSPPEELREIPDSEFERRVODMHHRRPDVAVYITTTGTARFLADDERVOLAVI	240
DB	181	ILALGSPPEELREIPDSEFERRVODMHHRRPDVAVYITTTGTARFLADDERVOLAVI	240
QY	241	GGGAGQLARLVGSPGHVFRHAECSVLVVR	272
DB	241	GGGAGQLARLVGSPGHVFRHAECSVLVVR	272

RESULT 2



```

US-10-617-038-21
: Sequence 21, Application US/10617038
: Publication NO. US20040057963A1
: GENERAL INFORMATION:
: APPLICANT: Andersen, Peter
: APPLICANT: Rosenkranz, Ida
: APPLICANT: Stryhn, Anette
: TITLE OF INVENTION: Therapeutic TB Vaccine
: FILE REFERENCE: SSI5AUSA
: CURRENT APPLICATION NUMBER: US/10/617,038
: CURRENT FILING DATE: 2003-07-11
: PRIOR APPLICATION NUMBER: DK PA 2002 01098
: PRIOR FILING DATE: 2002-07-13
: PRIOR APPLICATION NUMBER: US 60/401,725
: PRIOR FILING DATE: 2002-08-07
: NUMBER OF SEQ ID NOS: 187
: SOFTWARE: Patentin version 3.2
: SEQ ID NO 21
: LENGTH: 279
: TYPE: prt
: ORGANISM: Mycobacterium tuberculosis
US-10-617-038-21

```

```

Query Match          29.8%; Score 416; DB 4; Length 279;
Best Local Similarity 36.8%; Pred. No. 3e-31;
Matches 98; Conservative 46; Mismatches 116; Indels 6; Gaps 2;

QY      11 TTIVGIDGSHAATTAALMGVDEDAISRAPVLRIVSYIKETHPSPDDYD--RDLAHERSL 67
DB      9  SIVGIDDSKPAVQALMAWVEAASRDIPRLTYAIEPDDDEGYAHGAARKLAAMENAV 68
QY      68 REASAVAAKLVKIEETDIRGPAGPLVEASRAEMICVGVSGICGVYASSIIGSTATE 127
DB      69 RYATFAEAARPAKVEVEITQERPVTSLRISAAALVCVGAIGVHHFRPERGSTAA 128
QY      128 LAERAHCPVAWRSKVQOPASDINWIVRMTDAPDNEAVLEAARBAATROAPITALGCR 187
DB      129 LALSAQCVAIVRPHRPVIGRDAAIVVEADGSSDIGVLGAMVEARLRDSPRVVTCR 188
QY      188 PEELRELP--GEFERRVQDMHHRHPVRVYPTTTGIRFLADHDERVQAVIGGGE 244
DB      189 QSGVDGDDVRAISDRWLAMOPRYPDVRSAAVHGLDLYLAGLGRSVHMYVLSASD 248
QY      245 AGOLARLVGPSGHPVFERHAECSVLVV 270
DB      249 QEHVEQLVGAFGNAVLDZAGCTLLVV 274

RESULT 3
US-10-617-038-42
/ Sequence 42, Application US/10617038
/ Publication No. US20040057963A1
GENERAL INFORMATION:
/ APPLICANT: Andersen, Peter
/ APPLICANT: Rosenkrands, Ida
/ APPLICANT: Stryhn, Anette
/ TITLE OF INVENTION: Therapeutic TB Vaccine
/ FILE REFERENCE: SS15AUSA
CURRENT APPLICATION NUMBER: US/10/617,038
CURRENT FILING DATE: 2003-07-11
PRIOR APPLICATION NUMBER: DK PA 2002 01098
PRIOR FILING DATE: 2002-07-13
PRIOR APPLICATION NUMBER: US 60/401,725
PRIOR FILING DATE: 2002-08-07
NUMBER OF SEQ ID NOS: 187
SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 42
/ LENGTH: 268
/ TYPE: PRT
/ ORGANISM: Mycobacterium tuberculosis
US-10-617-038-42

25.6%; Score 357.5; DB 4; Length 268;

```

Best Local Similarity 36.5%; Pred. No. 1,28-25; Matches 100; Conservative 35; Mismatches 110; Indels 29; Gaps 6;

```

      RESULT 4          US-10-118-473-34
; Sequence 34, Application US/10138473
; Publication No. US20030165525X1
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: WELDINGH, Karln
; APPLICANT: HANSEN, Christlina Vegerby
; APPLICANT: FLORIO, Walter
; APPLICANT: OKKELS, Li Wei Meng
; APPLICANT: SKJORT, Rikke Louise Vinther
; APPLICANT: ROSENKRANDS, Ida
; APPLICANT: RASMUSSEN, Peter Birk
; TITLE OF INVENTION: TB diagnostic Based On Antigens From M. tuberculosis
; FILE REFERENCE: 0459-0710P
; CURRENT APPLICATION NUMBER: US/10/138,473
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US 10/060,428
; PRIOR FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: US 09/415,884
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 60/116,673
; PRIOR FILING DATE: 1999-01-21
; PRIOR APPLICATION NUMBER: DK 1998 01281
; PRIOR FILING DATE: 1998-10-18
; PRIOR APPLICATION NUMBER: US 60/070,488
; PRIOR FILING DATE: 1998-01-05
; PRIOR APPLICATION NUMBER: DK 1997 01277
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: US 60/044,624
; PRIOR FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: DK 1997 00376
; PRIOR FILING DATE: 1997-04-02
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-138-473-34

Query Match       19.1%; Score 267; DB 4; Length 297;
Best Local Similarity 29.3%; Pred.No. 6.6e-17;
Matches 84; Conservative 42; Mismatches 131; Indels 30; Gaps 8;

    Oy      12   IIVGIDSSHAATYALMGVDEAISRAVPRLRVSVIRK-----THPS-----DDYDR 58
            ||||| : | | | | : | | | | : | | | | : | | | | : | | | | : | | | |
Db         10   IIVGIDSPAAQAVRRARADAEIRKIPLTILVAVSFVAWTWLEVRPLPGVLRKQGDHR 69

```

```

Qy 59 DLAAERSLRBAQSAVEAAGKLVKIEITDIPRPAQPVLYEASRDAMI CVSGVIGRYAS 118
Db 70 HL--IDDLAKVVEQASLRAGP-PTVHSEIVPAAVPTLVDSKDAVLMVVCCLSGRMPG 126
Qy 119 SLIGSTATELAEKACHPVAWRSKYD-QPASDINMIVVMTDADNEAVLEYAAREAKLR 177
Db 127 RLSSVSSGLRHAHCPVVIITHDEDSVPHPOQAPVLGVGSSASSELATIADEASRR 186
Qy 178 CAPILALGRPE-ELREIP-----DGEFERRVQDWHHHPDVVVYPITTHGIA 225
Db 187 NVDLVALHAMSDDVSEVMPGIDWPATQSMAGVLAERLAGQGERYPVNAITRVVVRQPA 246
Qy 226 RFLADHDERVOLAVIGG-GEAGQLARLVGSGHVFRRHAECSVLVR 271
Db 247 ROLVORSEBAQLVVVSGRGGYAGMLVGSVGETVAQIARFPVIVAR 293

```

## RESULT 5

```

US-10-450-726-4
; Sequence 4, Application US/10450726
; Publication No. US20040242471A1
; GENERAL INFORMATION:
; APPLICANT: Dick, Thomas
; TITLE OF INVENTION: DORMANCY-INDUCED MYCOBACTERIUM PROTEINS
; FILE REFERENCE: 50318/002001
; CURRENT APPLICATION NUMBER: US/10/450,726
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: PCT/EP01/14551
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 0030368.5
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-450-726-4

```

```

Query Match 19.1%; Score 267; DB 5; Length 297;
Best Local Similarity 29.3%; Pred. No. 6.6e-17;
Matches 84; Conservative 42; Mismatches 131; Indels 30; Gaps 8;

Qy 12 IIVGIDGSHAATAITAAWGVDEAISRAVPLRLVSYIKP-----THPSP-----DDYDR 58
Db 10 IIVGIDGSPAQAQVAVRAAARDAERKIFLITVHAHVSEVATWLEVPPLPCVLTLMQDQHR 69
Qy 59 DLAAERSLRBAQSAVEAAGKLVKIEITDIPRPAQPVLYEASRDAMI CVSGVIGRYAS 118
Db 70 HL--IDDLAKVVEQASLRAGP-PTVHSEIVPAAVPTLVDSKDAVLMVVCCLSGRMPG 126
Qy 119 SLIGSTATELAEKACHPVAWRSKYD-QPASDINMIVVMTDADNEAVLEYAAREAKLR 177
Db 127 RLSSVSSGLRHAHCPVVIITHDEDSVPHPOQAPVLGVGSSASSELATIADEASRR 186
Qy 178 CAPILALGRPE-ELREIP-----DGEFERRVQDWHHHPDVVVYPITTHGIA 225
Db 187 NVDLVALHAMSDDVSEVMPGIDWPATQSMAGVLAERLAGQGERYPVNAITRVVVRQPA 246
Qy 226 RFLADHDERVOLAVIGG-GEAGQLARLVGSGHVFRRHAECSVLVR 271
Db 247 ROLVORSEBAQLVVVSGRGGYAGMLVGSVGETVAQIARFPVIVAR 293

```

## RESULT 6

```

US-09-738-626-6745
; Sequence 6745, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO

```

```

; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 6745
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6745

```

```

Query Match 19.0%; Score 266; DB 3; Length 301;
Best Local Similarity 28.8%; Pred. No. 8.4e-17;
Matches 85; Conservative 45; Mismatches 131; Indels 34; Gaps 8;

Qy 8 TMTIIVGIDGSHAATAITAAWGVDEAISRAVPLRLVSYIKP-----PDYDR 58
Db 3 TEDIIVAAVDGSPASKCAVMAANTANKGIPRLAS--SYTPOFLYAGWVPPQFLFD 60
Qy 59 DLAAERSLRBAQSAVEAAGKLVKIEITDIPRPAQPVLYEASRDAMI CVSGVIGRYA 117
Db 61 DLQAEALKEKINARDIAHEPAPEIKGHTAESSPIDMLEMSPDAMTMYGSRGLGLS 120
Qy 118 SSLIGSTATELAEKACHPVAWNR--SKVDOPASDINMIVVMTDADNEAVLEYAAREAK 175
Db 121 GMMWGSVGAIVSHAKCFVVVVEDSAVNED-SKYGPVVVGVGSEVSQATEYAFEAEE 179
Qy 176 LR-----QAPILALGRPEELRIIPDGEFE---RRVQDWHHHPDVVVYP 217
Db 180 ARGAEIYAVHTWMDMQVQSLAGIAAQQOWDEVERQOTDMLERLAPVEKYPSTVVKX 239
Qy 218 ITTHGTARFLADHDERVOLAVIGG-GEAGQLARLVGSGHVFRRHAECSVLVR 271
Db 240 IITRDPRVALAASENAQLVVVSGHGGFKGMLGSTRALLQSAFCPMVVR 294

```

## RESULT 7

```

US-09-746-660A-6
; Sequence 6, Application US/09746660A
; Publication No. US2003049804A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Krogger, Burhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Habermeyer, Gregor
; APPLICANT: Kim, Jun-won
; APPLICANT: Lee, Heung-Schick
; APPLICANT: Hwang, Byung-Joon
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING METABOLIC PATHWAY PROTEINS
; FILE REFERENCE: BGI-121CP2
; CURRENT APPLICATION NUMBER: US/09/746,660A
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 09/606740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 09/603124
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141031

```



Db 240 IITDRPVRALAEASNAQLVVGSHRGSGFKMGLGSTSRALLQSAFCPMVVVR 294

RESULT 10

US-11-006-098-314  
 ; Sequence 314, Application US/11006098  
 ; Publication No. US20050153402A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Pompejus, Markus  
 ; APPLICANT: Kroger, Burkhard  
 ; APPLICANT: Schröder, Hartwig  
 ; APPLICANT: Zelder, Oskar  
 ; APPLICANT: Haberhauser, Gregor  
 ; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
 ; TITLE OF INVENTION: REGULATORY PROTEINS  
 ; FILE REFERENCE: BGI-123CP  
 ; CURRENT APPLICATION NUMBER: US/11/006,098  
 ; CURRENT FILING DATE: 2004-12-06  
 ; PRIOR APPLICATION NUMBER: US/09/602,874  
 ; PRIOR FILING DATE: 2000-06-23  
 ; PRIOR APPLICATION NUMBER: 60/141031  
 ; PRIOR FILING DATE: 1999-06-25  
 ; PRIOR APPLICATION NUMBER: 60/142690  
 ; PRIOR FILING DATE: 1999-07-01  
 ; PRIOR APPLICATION NUMBER: 60/151251  
 ; PRIOR FILING DATE: 1999-08-27  
 ; PRIOR APPLICATION NUMBER: DE 19930476.9  
 ; PRIOR FILING DATE: 1999-07-01  
 ; PRIOR APPLICATION NUMBER: DE 19931419.5  
 ; PRIOR FILING DATE: 1999-07-07  
 ; PRIOR APPLICATION NUMBER: DE 19931420.9  
 ; PRIOR FILING DATE: 1999-07-08  
 ; PRIOR APPLICATION NUMBER: DE 19932122.1  
 ; PRIOR FILING DATE: 1999-07-09  
 ; PRIOR APPLICATION NUMBER: DE 19932128  
 ; PRIOR FILING DATE: 1999-07-09  
 ; PRIOR APPLICATION NUMBER: DE 19932206.6  
 ; PRIOR FILING DATE: 1999-07-09  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 362  
 ; SEQ ID NO 314  
 ; LENGTH: 301  
 ; TYPE: PRT  
 ; ORGANISM: Corynebacterium glutamicum  
 ; US-11-006-098-314

Query Match 19.0%; Score 266; DB 6; Length 301;  
 Best Local Similarity 28.8%; Pred. No. 8.4e-17;  
 Matches 85; Conservative 45; Mismatches 131; Indels 34; Gaps 8;

QY 8 TMTIIVIGDSNAATPAALMGVDEAISRAVPLRLVSVIKRTHSPD-----PDYDR 58  
 DB 3 TEDIIVAVVSDSDSKQAVRWAAANTANKRGIPRLIAS--SYTMPQFLVYAEQWVPQELFDEL 60  
 QY 59 DL-AHARSRSREASAVEAAGKLVKTIENDIPRGAPVLVEASRDAEMICVSGVIGRKYA 117  
 DB 61 DLQALAEKINEADIDIAHEVAPEIKIGTTIAEGSPIDMLDMSPDATYIVWGSRGIGLS 120  
 QY 118 SSIIIGSTATELEAKHCPVAVNR--SKVDPASDINMIVAMTDA PDNEAVLEVAABEAK 175  
 DB 121 GMMVGSVGAIVSHAKCPVVVVVREDSANVED-SKYGPPVVGVDSDSVSQATEYAFABAE 179  
 QY 176 LR-----CAPILAGSRPEELREIPDGEF--RRVQDMHHRHDPVRYVP 217  
 DB 180 ARGAEIVAVHTMDMOVQASLAGLAAQOQWDEVERQDTDLIRLAPLVYKESYVYVKK 239  
 QY 218 ITHHTGATPLADDERVQGLAVIGG-GEAGQLARLVGSPSHVFRHAEGSVLYVR 271  
 DB 240 IITDRPVRALAEASNAQLVVGSHRGSGFKMGLGSTSRALLQSAFCPMVVVR 294

RESULT 11

US-10-282-122A-53861  
 ; Sequence 53861, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELITRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10/282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 53861  
 ; LENGTH: 300  
 ; TYPE: PRT  
 ; ORGANISM: Corynebacterium diptheriae  
 ; US-10-282-122A-53861

Query Match 17.8%; Score 249; DB 4; Length 300;  
 Best Local Similarity 28.5%; Pred. No. 3.6e-15;  
 Matches 84; Conservative 46; Mismatches 123; Indels 42; Gaps 10;

QY 12 IIVGIDGSHAAITPAALMGVDEAISRAVPLRLVSVIKRTHSPD--YDRDLAHAERSIREA 70  
 DB 7 VVAVVSDSDSKQAVRWAAANTANKRGVPLRLIAA-----SYTMPQFLVYAEQWVPQELFDEL 62  
 QY 71 QS-----AVEAAGKLVKTIENDIPRGAPVLVEASRDAEMICVSGVIGRKYA 119  
 DB 63 QSETMDIIEARVAHEVADIKIGYVIAEGSPIDMLDMSPDVMTIWMGSRGIGLSGM 122  
 QY 120 ILGSTATELEAKHCPVAVNRSKVDPASDINM--IIVAMTDA PDNEAVLEVAABEAKL 176  
 DB 123 VMGSVAIVSHADCPVVVVR--DHNVTETNKKGPPVVGVDSDSVSQATEYAFABEAAQ 180  
 QY 177 R-----CAPILAGSRPEELREIPDGE--RRVQDMHHRHDPVRYVP 217  
 DB 181 RGAEIVAVHTMDMOVQASLAGLAAQOEW-EIIEKEQFTLLKDRLOPLERPPDVEVEM 239  
 QY 218 ITHHTGATPLADDERVQGLAVIGG-GEAGQLARLVGSPSHVFRHAEGSVLYVR 271  
 DB 240 VITDRPVRALAECAHNAQLLVVGSHRGSGFKMGLGSTSRALLQSAFCPMVVVR 294





This page Blank (uspto)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Biocelelation Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 06:17:23 ; Search time 6.52731 Seconds  
(without alignments)  
1192.746 Million cell updates/sec

Title: US-10-617-038-27

Perfect score: 1397

Sequence: 1 MSGGEPMTTIVIGDSH.....GPSGHPVFRHAECSVLVRR 272

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 169630 seqs, 2862289 residues

Total number of hits satisfying chosen parameters: 169630

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	266	19.0	301	US-11-055-822-14	Sequence 14, Appl
2	262.5	18.8	305	US-11-087-099-3468	Sequence 3468, Ap
3	149	10.7	152	US-11-087-099-6347	Sequence 6347, Ap
4	130.5	9.3	163	US-11-087-099-11409	Sequence 11409, A
5	130.5	9.3	163	US-11-096-568A-4791	Sequence 4791, Ap
6	130.5	9.3	165	US-11-096-568A-4790	Sequence 4790, Ap
7	128.5	9.2	171	US-11-087-099-505	Sequence 505, App
8	128.5	9.2	171	US-11-096-568A-15853	Sequence 15853, A
9	125.5	9.0	153	US-11-096-568A-15854	Sequence 15854, A
10	122.5	8.8	174	US-11-096-568A-27471	Sequence 27471, A
11	122	8.7	171	US-11-096-568A-32572	Sequence 32572, A
12	122	8.7	179	US-11-096-568A-27882	Sequence 27882, A
13	120.5	8.6	151	US-11-096-568A-15855	Sequence 15855, A
14	120	8.6	172	US-11-087-099-11256	Sequence 11256, A
15	118.5	8.5	175	US-11-087-099-7223	Sequence 7223, Ap
16	117.5	8.4	152	US-11-096-568A-4792	Sequence 4792, Ap
17	116.5	8.3	162	US-11-087-099-436	Sequence 436, App
18	116.5	8.3	169	US-11-087-099-2259	Sequence 2259, Ap
19	116	8.3	169	US-11-087-099-10381	Sequence 10381, Ap
20	114	8.2	169	US-11-087-099-4994	Sequence 4994, Ap
21	112	8.0	182	US-11-087-099-2660	Sequence 2660, Ap
22	109	7.8	186	US-11-087-099-7765	Sequence 7765, Ap
23	107.5	7.7	150	US-11-087-099-3411	Sequence 3411, Ap
24	107	7.7	175	US-11-087-099-9020	Sequence 9020, Ap
25	106.5	7.6	162	US-11-096-568A-1508	Sequence 1508, Ap

26	106.5	7.6	181	US-11-087-099-10726	Sequence 10726, A
27	103.5	7.4	162	US-11-087-099-5731	Sequence 5731, Ap
28	103	7.4	159	US-11-087-099-12270	Sequence 12270, A
29	103	7.4	161	US-11-087-099-2546	Sequence 2546, Ap
30	101.5	7.3	162	US-11-087-099-9120	Sequence 9120, Ap
31	101.5	7.3	162	US-11-087-099-10134	Sequence 10134, A
32	101.5	7.3	203	US-11-096-568A-10261	Sequence 10261, A
33	101	7.2	153	US-11-096-568A-1509	Sequence 1509, Ap
34	100.5	7.2	160	US-11-087-099-3178	Sequence 3178, Ap
35	99	7.1	125	US-11-087-099-6096	Sequence 6096, Ap
36	99	7.1	161	US-11-087-099-6989	Sequence 6989, Ap
37	99	7.1	162	US-11-096-568A-10262	Sequence 10262, A
38	99	7.1	171	US-11-087-099-8466	Sequence 8466, Ap
39	99	7.1	171	US-11-087-099-8499	Sequence 8499, Ap
40	98.5	7.1	270	US-10-467-657-1630	Sequence 1630, Ap
41	98	7.0	171	US-11-087-099-7706	Sequence 7706, Ap
42	97.5	7.0	164	US-11-087-099-6603	Sequence 6603, Ap
43	97.5	7.0	3655	US-11-075-185-5	Sequence 5, Appl
44	97	6.9	154	US-11-096-568A-27472	Sequence 27472, A
45	97	6.9	179	US-11-096-568A-6711	Sequence 6711, Ap

## ALIGNMENTS

RESULT 1  
US-11-055-822-14  
; Sequence 14, Application US/11055822  
; Publication NO. US20050260707A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompeius, Markus  
; APPLICANT: Krogger, Burkhard  
; APPLICANT: Schröder, Hartwig  
; APPLICANT: Zeidler, Oskar  
; APPLICANT: Haberhauser, Gregor  
; TITLE OF INVENTION: CORINEBACTERIUM GLUTAMICUM GENES ENCODING  
; FILE REFERENCE: BGI-121CPCN  
; CURRENT APPLICATION NUMBER: US/11/055, 822  
; PRIOR FILING DATE: 2005-02-11  
; PRIOR APPLICATION NUMBER: 09/606,740  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 60/141,031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: 60/142,101  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: 60/148,613  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/187,970  
; PRIOR FILING DATE: 2000-03-09  
; PRIOR APPLICATION NUMBER: DE 19930476.9  
; PRIOR FILING DATE: 1999-07-01  
; PRIOR APPLICATION NUMBER: DE 19931415.2  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19931418.7  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19931419.5  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19931420.9  
; PRIOR FILING DATE: 1999-07-08  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1158  
; SEQ ID NO 14  
; LENGTH: 301  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
; US-11-055-822-14

Query Match 19.0%; Score 266; DB 7; Length 301;  
Best Local Similarity 28.8%; Pred. No. 1,1e-15;  
Matches 85; Conservative 45; Mismatches 131; Indels 34; Gaps 8;  
Qy 8 TMKTIIVIGDSHAATPALMGVDEAISRAVPLRLVSVIKRTHPS-----PDDYDR 58



Db 3 TEDIIVAAVDSGDSKQAVRNAATANKRGIPRLAAS--SYTMQFLYAEQMPVQBEFD 60  
Qy 59 DL-AHAERSLREASAVEAAGKLVKIETDIPRGAPVLYEASRDAMI CVGSGVIGRYA 117  
Db 61 DLQALEKINEADIDIAHEVAPEIKGHTIAEGSPIDMLMSPDATMI VVSGRLGGLS 120  
Qy 118 SSILGSTATLAEAKHCPVAVMR--SKYDOPASDINMIVRMTPDA PDNEAVLEYAAREAK 175  
Db 121 GVVVGSVGA VVSHAKCPVVVVVRRSDAVNED--SKYGPVVGVDSGEVSSQATLEYAFAEAE 179  
Qy 176 LR-----CAPILALGRRPELEIRPDGEFE---RRVQDMHHRHDPVAVR 217  
Db 180 ARGAEIIVAHVTMDMVOVASLGLAAAOQOWDEVERQOTDMLIRLAPLYEKYPSVTVVK 239  
Qy 218 ITTHTGJARFLADDERVOLAVIGG-GEAQLARLVGSPGHVPRHAECVLYVR 271  
Db 240 IITDRPRVALAESENAQLLVGSHKRGFKGMLLSTSRALLQSAPCPMAYVR 294

## RESULT 2

US-11-087-099-3468  
; Sequence 3468, Application US/11087099  
; Publication No. US20060041961A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53450) B EP  
; CURRENT APPLICATION NUMBER: US/11/087,099  
; CURRENT FILING DATE: 2005-03-22  
; NUMBER OF SEQ ID NOS: 12464  
; SEQ ID NO 3468  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Thermodifida fusca  
US-11-087-099-3468

Query Match 18.8%; Score 262.5; DB 7; Length 305;  
Best Local Similarity 28.7%; Pred. No. 2.3e-15;  
Matches 89; Conservative 41; Mismatches 131; Indels 49; Gaps 8;

Qy 1 MSGGGEPTMTTIYIGIDGSHAITAALMGVDEAISRAVPLRLV-----SYTK----- 48  
Db 1 MSBR-ETVPGTVVVGVDSBGSILHALDMAIDAAAGRGATRLVYAMGLPLTVTLVGGPIR 59  
Qy 49 THSPDDYDRLDLAHERSLREASAVEAAGKLVKIETDIPRGAPVLYEASRDAMICV 108  
Db 60 TABSPVSOAKKALLEALRVRVQSAFS---LRATVEVSAEHAHALLISAQDAELIIV 115  
Qy 109 GSVGIGRYASSIIGSTATLAEAKHCPVAVMRSKVDOPASDINMIVRMTPDAPNEAVLE 168  
Db 116 GSRGYSVGA SLFGSVAGRVASHATCPVVVVPPTSGEAAARGRVVGVDSGEHAAALR 175  
Qy 169 YAAREALRQAPILALGRRPELEIRPDGEFE-----RRVQDMHHRP 211  
Db 176 FALVEARLRALRELVAV---YAWQAPDAPVDPPTVLQADVADREOVAAAREMLRLTV 230  
Qy 212 DVRYVPTTTTGT-----ARFLADDERVOLAVIGG-GEAQLARLVGSPGHVPR 261  
Db 231 DEARTPLTORPVAVVERPEKHPAALLDEGADALIVGSGRGGFTGLIGSVSQVYN 290  
Qy 262 HAECVLYVR 271  
Db 291 HAVPVAVVR 300

## RESULT 3

US-11-087-099-6347  
; Sequence 6347, Application US/11087099  
; Publication No. US20060041961A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement

; FILE REFERENCE: 38-21(53450) B EP  
; CURRENT APPLICATION NUMBER: US/11/087,099  
; CURRENT FILING DATE: 2005-03-22  
; NUMBER OF SEQ ID NOS: 12464  
; SEQ ID NO 6347  
; LENGTH: 152  
; TYPE: PRT  
; ORGANISM: Streptomyces coelicolor A3(2)  
US-11-087-099-6347

Query Match 10.7%; Score 149; DB 7; Length 152;  
Best Local Similarity 29.3%; Pred. No. 6e-06;  
Matches 44; Conservative 26; Mismatches 62; Indels 18; Gaps 5;

Qy 6 EPTMTTIYIGIDGSHAITAALMGVDEAISRAVPLRLVSYKPTH-----PS-----PD 54  
Db 5 EPTAR-VVGVDSPPSYALRNA--DRYARAVG---GVVEAVHWDTPTSAGVAPAGAI 57  
Qy 55 DYDRDLAHERSLREASAVEAAGKLVKIETDIPRGAPVLYEASRDAMICVSGVIG 114  
Db 58 DPDFDLEQARERPAELEATFPGERPGLKEIIVEGDPSETLIRASGAEIIVGRRGRG 117  
Qy 115 RVASSIIGSTATLAEAKHCPVAVMRSKVD 144  
Db 118 AFARAMLGVSQRCAQHACPVVVRQETE 147

## RESULT 4

US-11-087-099-11409  
; Sequence 11409, Application US/11087099  
; Publication No. US20060041961A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53450) B EP  
; CURRENT APPLICATION NUMBER: US/11/087,099  
; CURRENT FILING DATE: 2005-03-22  
; NUMBER OF SEQ ID NOS: 12464  
; SEQ ID NO 11409  
; LENGTH: 163  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-11-087-099-11409

Query Match 9.3%; Score 130.5; DB 7; Length 163;  
Best Local Similarity 29.3%; Pred. No. 0.00026;  
Matches 46; Conservative 25; Mismatches 69; Indels 17; Gaps 6;

Qy 1 MSGRGEPTMTTIYIGIDGSHAITAALMGVDEAISRA-VPLRLVSY--KPTSPDDY- 56  
Db 1 MATSGSETKQVWVIGIDSDPSFYALQWTLDLHLSPANVPKFKIPLVYAKPSVASAVGV 60  
Qy 57 ----DRDLAHERSLRE-AQSAVEAAGKLVK-----IETDIPRGAPVLYEA--SRDA 103  
Db 61 GPAAAEVLPVVEADLRTAKITERATBELCKKSVNDVAVEVLEGDPNNVLCVAVEKQA 120  
Qy 104 EMICVSGVIGRYASSIIGSTATLAEAKHCPVAVMR 140  
Db 121 SMLVGSHGVTGLKRAVIGSVSDYCAHHAHCTMAYK 157

## RESULT 5

US-11-096-568A-4791  
; Sequence 4791, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471

```
; SEQ ID NO 4791
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(163)
; OTHER INFORMATION: Ceres Seq. ID no. 14304185
US-11-096-568A-4791
```

```
Query Match          9.3%; Score 130.5; DB 7; Length 163;
Best Local Similarity 29.3%; Pred. No. 0.00026;
Matches 46; Conservative 25; Mismatches 69; Indels 17; Gaps 6;
```

```
QY 1 MSGRPEPTMTITIVIGIDGSHAITTAALMGVDEAISR-A-VPLRLVSVI--KPTHPSPDDY- 56
   1 MATSSSETKQVWVIGIDSDSFSTYALQWTLDHLLSPAVVPKIKILVYAKSVASAVGFV 60
DB 57 ----DRDLAHERSLR-AGSAVEAAGKLVK-----IETDIPRGPAPVLEA--SRDA 103
   61 GPGAAEVLPPVEADLRKTAAPKITERATELCCKKSVNDVAVEVLLEGDPNRVLCBAVEKIQ 120
QY 104 EMICVSGVIGIRYASSIIIGSTATELAERAKHCPVAVMR 140
   121 SMLVVGSHGIGTLKRAVLGVSVDYCAHHHACTVMTIVK 157
DB
```

```
RESULT 6
US-11-096-568A-4790
; Sequence 4790, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Theryby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 4790
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(165)
; OTHER INFORMATION: Ceres Seq. ID no. 14304184
US-11-096-568A-4790
```

```
Query Match          9.3%; Score 130.5; DB 7; Length 165;
Best Local Similarity 29.3%; Pred. No. 0.00027;
Matches 46; Conservative 25; Mismatches 69; Indels 17; Gaps 6;
```

```
QY 1 MSGRPEPTMTITIVIGIDGSHAITTAALMGVDEAISR-A-VPLRLVSVI--KPTHPSPDDY- 56
   1 MATSSSETKQVWVIGIDSDSFSTYALQWTLDHLLSPAVVPKIKILVYAKSVASAVGFV 62
DB 3 MATSSSETKQVWVIGIDSDSFSTYALQWTLDHLLSPAVVPKIKILVYAKSVASAVGFV 62
QY 57 ----DRDLAHERSLR-AGSAVEAAGKLVK-----IETDIPRGPAPVLEA--SRDA 103
   63 GPGAAEVLPPVEADLRKTAAPKITERATELCCKKSVNDVAVEVLLEGDPNRVLCBAVEKIQ 122
DB 63 GPGAAEVLPPVEADLRKTAAPKITERATELCCKKSVNDVAVEVLLEGDPNRVLCBAVEKIQ 122
QY 104 EMICVSGVIGIRYASSIIIGSTATELAERAKHCPVAVMR 140
   123 SMLVVGSHGIGTLKRAVLGVSVDYCAHHHACTVMTIVK 159
DB
```

```
RESULT 7
US-11-087-099-505
; Sequence 505, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abadi, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
```

```
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 505
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Zea mays
US-11-087-099-505
```

```
Query Match          9.2%; Score 128.5; DB 7; Length 171;
Best Local Similarity 34.0%; Pred. No. 0.00042;
Matches 51; Conservative 14; Mismatches 68; Indels 17; Gaps 7;
```

```
QY 7 PTMKTIIVIGIDGSHAITTAALMGVDEAISR-AP--LRLVSV-IRKTHPS-----PDDYD 57
   16 PGKMTWVGVDESHSFPALQWALQHPPRPPQYRLVVVTAKPTAASAVGLAGPGAAD 75
DB 58 RDLAHERSL-REAGSAVEAAGKLVKJETDIP----RGPAPVLEA--SRDAEMICVGS 110
   76 -VLPVEADLRKRSALRVEERAKGLCTQASDAVFPEALBGDARNVLCBAVERHGAEMLVGS 134
QY 111 VGIGRYASSIIIGSTATELAERAKHCPVAVMR 140
   135 HGYGAIKRAVLGVSVDYCAHHHACTVMTIVK 164
DB
```

```
RESULT 8
US-11-096-568A-15853
; Sequence 15853, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Theryby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 15853
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(171)
; OTHER INFORMATION: Ceres Seq. ID no. 12349153
US-11-096-568A-15853
```

```
Query Match          9.2%; Score 128.5; DB 7; Length 171;
Best Local Similarity 34.0%; Pred. No. 0.00042;
Matches 51; Conservative 14; Mismatches 68; Indels 17; Gaps 7;
```

```
QY 7 PTMKTIIVIGIDGSHAITTAALMGVDEAISR-AP--LRLVSV-IRKTHPS-----PDDYD 57
   16 PGKMTWVGVDESHSFPALQWALQHPPRPPQYRLVVVTAKPTAASAVGLAGPGAAD 75
DB 16 PGKMTWVGVDESHSFPALQWALQHPPRPPQYRLVVVTAKPTAASAVGLAGPGAAD 75
QY 58 RDLAHERSL-REAGSAVEAAGKLVKJETDIP----RGPAPVLEA--SRDAEMICVGS 110
   76 -VLPVEADLRKRSALRVEERAKGLCTQASDAVFPEALBGDARNVLCBAVERHGAEMLVGS 134
DB 76 -VLPVEADLRKRSALRVEERAKGLCTQASDAVFPEALBGDARNVLCBAVERHGAEMLVGS 134
QY 111 VGIGRYASSIIIGSTATELAERAKHCPVAVMR 140
   135 HGYGAIKRAVLGVSVDYCAHHHACTVMTIVK 164
DB
```

```
RESULT 9
US-11-096-568A-15854
; Sequence 15854, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-determined DNA Fragments and Corresponding Polypeptides
```

```

; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 15854
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(153)
; OTHER INFORMATION: Cereb Seq. ID no. 12349154
US-11-096-568A-15854

Query Match      9.0%; Score 125.5; DB 7; Length 153;
Best Local Similarity 34.2%; Pred. No. 0.00066;
Matches 50; Conservative 14; Mismatches 65; Indels 17; Gaps 7;

QY 11 TTVIGIDGSHAAITPAALMGVDEAISRVP--LRIVSV-IKPTSPD-----PDDYDRDLA 61
DB 2 TMVVGVESEHSFPALQWALOHFFPPGQPOQYRLVVTAKFTPAASAVGLAIPGAD-VLP 60
QY 62 HAERSL-REAGSAVEAAGKLVKIETDIP---RGPAQVLYEA--SRDAEMICVSGVIG 114
DB 61 YVEADLRKRSALRVVEKAGLCTQASDAVFALBEGDANNVLCBAVERHGAEMLVVSGHYG 120
QY 115 RYASSIIIGSTATELAERKACPAVAVR 140
DB 121 AIKRAVLGSVSDYCAHHNHCVTMTVK 146

RESULT 10
US-11-096-568A-27471
; Sequence 27471, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 27471
; LENGTH: 174
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(174)
; OTHER INFORMATION: Cereb Seq. ID no. 1817050
US-11-096-568A-27471

Query Match      8.8%; Score 122.5; DB 7; Length 174;
Best Local Similarity 22.5%; Pred. No. 0.0014;
Matches 41; Conservative 25; Mismatches 61; Indels 55; Gaps 5;

QY 8 TMKTIIVGIDGSHAAITPAALMGVDEAISRVPRLVS-----VIKPTSPD----- 54
DB 6 SLNCVVVAVDGSEVSMELRMALDN-----LKLSSSSSSSSSFVLVHQPSPSVAAVS 58
QY 55 -----DYDRDLHAERSLREAGSAVEAAGKLVKIETDI 87
DB 59 PGTIFGPGSGLEVPAPFTAAIEHQKRIPTLIEHSA-----SQICAEKSVSRVNVKTQV 112
QY 88 PRGPAGVLYEASRD--AEMICVSGVIGRYASSIIGSTATELAERKACHPAVAMRSKVQ 145
DB 113 VIDPPTYKICEAVENIHADLVWGSRAYGRIKRMFLGVSVNYCTNHAHCPVITIKPEDS 172
QY 146 PA 147
DB 173 SA 174
```

```

RESULT 11
US-11-096-568A-32572
; Sequence 32572, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 32572
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(171)
; OTHER INFORMATION: Cereb Seq. ID no. 13593512
US-11-096-568A-32572

Query Match      8.7%; Score 122; DB 7; Length 171;
Best Local Similarity 22.5%; Pred. No. 0.0015;
Matches 41; Conservative 24; Mismatches 59; Indels 58; Gaps 5;

QY 8 TMKTIIVGIDGSHAAITPAALMGVDEAISRVPRLVS-----VIKPTSPD----- 54
DB 6 SLNCVVVAVDGSEVSMELRMALDN-----LKLSSSSSSSSSFVLVHQPSPSVAAVS 58
QY 55 -----DYDRDLHAERSLREAGSAVEAAGKLVKIETDI 87
DB 59 PGTIFGPGSGLEVPAPFTAAIEHQKRIPTLIEHSAQCAB-----KSNVKTQV 109
QY 88 PRGPAGVLYEASRD--AEMICVSGVIGRYASSIIGSTATELAERKACHPAVAMRSKVQ 145
DB 110 VIDPPTYKICEAVENIHADLVWGSRAYGRIKRMFLGVSVNYCTNHAHCPVITIKPEDS 169
QY 146 PA 147
DB 170 SA 171

RESULT 12
US-11-096-568A-27882
; Sequence 27882, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 27882
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(179)
; OTHER INFORMATION: Cereb Seq. ID no. 2123664
US-11-096-568A-27882

Query Match      8.7%; Score 122; DB 7; Length 179;
Best Local Similarity 23.2%; Pred. No. 0.0016;
Matches 43; Conservative 25; Mismatches 61; Indels 56; Gaps 6;

QY 8 TMKTIIVGIDGSHAAITPAALMGVDEAISRVPRLVS-----VIKPTSPD----- 54
DB 173 SA 174
```

```

Db      6 SLNCVVAVDGSEVSEALRNALDN-----LKLSSSSSDSSFVVLHVQSPSVAAGVS 58
Qy      55 -----DYDRDLAH-----AERSLREAGSAVEAAGLVKIE 84
Db      59 PGTFPGGGLVEPAPTAIEOHOKRITDTILEHSAQICAEKSVSRVLSLSP-----VAVK 114
Qy      85 TDIPRGPAPVLVEASRD--AEMI CVGSGVIGRYASSILGSTATELAERKACPVAVMRSK 142
Db      115 TQVAVGPBKVKICAEVENTLHADLLVMGSRAYRIKRMPLGSSVSVCTHNAHCPVYIIRKX 174
Qy      143 VDQPA 147
Db      175 EDSSA 179

```

## RESULT 13

```

US-11-096-568A-15855
; Sequence 15855, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 15855
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(151)
; OTHER INFORMATION: Ceres Seq. ID no. 12349155
US-11-096-568A-15855

```

Query Match 8.6%; Score 120.5; DB 7; Length 151;  
 Best Local Similarity 33.8%; Pred. No. 0.0017;  
 Matches 49; Conservative 14; Mismatches 65; Indels 17; Gaps 7;

```

Qy      12 IIVGIDGSHAATTAALMGVDEAISRVP--LRIVSV-IKPTHPG-----PDDYDRDLAH 62
Db      1 MIVGVDESEHSFYALQWALQHFPFGQEQRYRLVVVTKAPPAASAVGLAGGAD-VLPY 59
Qy      63 AERSL-REAGSAVEAAGLVKIEITDIP-----RGPAGPVLVEA--SRDAEMICVSGVIGR 115
Db      60 VEAULKRSALRVEKAKGLCTQASDAVEALEGDAARNVLCFAVERHGAEMLVGSHGYGA 119
Qy      116 YASIIIGSTATELAERKACPVAVMR 140
Db      120 IKRAVLGSSVDYCAHNAHCTWIVK 144

```

## RESULT 14

```

US-11-087-099-11256
; Sequence 11256, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 11256
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Trifolium aestivum
US-11-087-099-11256

```

Query Match 8.6%; Score 120; DB 7; Length 172;  
 Best Local Similarity 31.0%; Pred. No. 0.0023;

```

Matches 48; Conservative 19; Mismatches 66; Indels 22; Gaps 7;
Qy      3 GRGEPTKTIIVGIDGSHAATTAALMGVDEAIS--RAVPLRIVSV-IKPT-----HP 51
Db      17 GEGKMTV---VGVDESEHSFYALQWTLHFPPGQOQRYRLVVVTKAPPAASAVGLAGP 73
Qy      52 SPDD---YDRDLAHAERSLREAGSAVEAAGLVKIEITDIPRGPAGPVLVEA--SRDAEM 105
Db      74 GAADVLPFVEADLKRS--SLRVIDKAKELCAQVSDAVEFEVGGDARNVLCFAVERHHAEM 131
Qy      106 ICVGSVIGRYASSILGSTATELAERKACPVAVMR 140
Db      132 LVVGNHGYGAIKRAVLGSSVDYCTHNAHCTWIVK 166

```

## RESULT 15

```

US-11-087-099-7223
; Sequence 7223, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 7223
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(175)
; OTHER INFORMATION: unsure at all Xaa locations
US-11-087-099-7223

```

Query Match 8.5%; Score 118.5; DB 7; Length 175;  
 Best Local Similarity 26.0%; Pred. No. 0.0032;  
 Matches 45; Conservative 15; Mismatches 54; Indels 59; Gaps 5;

```

Qy      7 PYMKTIIVGIDGSHAATTAALMGVDEAI---SRAVPLRIVSVIKPTHPSPDDYDRDLAH 62
Db      16 PGKRTVVGVDESEHSFYALQWALQHFPFGAPPAANPPSSCVTAKPX----- 62
Qy      63 AERSLREAGSAVEAAGK-----LVKIEITDIP----- 89
Db      63 -----AASAVGLAGPXAADVLPYVEADLKRSALRVEKAKGLCTQVILASDAVEALE 115
Qy      90 GPAGPVLVEA--SRDAEMICVSGVIGRYASSILGSTATELAERKACPVAVMR 140
Db      116 GDARNVLCFAVERHGAEMLVGSHGYGAIKRAVLGSSVDYCAHNAHCTWIVK 168

```

Search completed: March 23, 2006, 06:29:37  
 Job time : 6.52731 secs

This page Blank (uspto)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

## OM protein - protein search, using sw model

Run on: March 23, 2006, 04:56:29 ; Search time 104.486 Seconds  
(without alignments)  
1652.628 Million cell updates/sec

Title: US-10-617-038-28

Perfect score: 1990  
Sequence: 1 MRDIPILGRAGFVVNVHMS.....QPEPFTTSPQDADRPSDAG 393

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

A\_Geneseq\_21:.\*  
1: geneseqp1980s:.\*  
2: geneseqp1980s:.\*  
3: geneseqp2000s:.\*  
4: geneseqp2001s:.\*  
5: geneseqp2002s:.\*  
6: geneseqp2003as:.\*  
7: geneseqp2003bs:.\*  
8: geneseqp2004s:.\*  
9: geneseqp2005s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	1990	100.0	393	8	AD137307 M. tuberc
2	425.5	21.4	379	5	ABG77902 High leve
3	244	12.3	290	5	ABM96841 M. xanthu
4	174	8.7	346	9	ABM94210 M. xanthu
5	154.5	7.8	438	8	ADN48111 Thermococ
6	143	7.2	409	4	AAB96393 Putative
7	142.5	7.2	177	7	ADM26447 Hyperther
8	140	7.0	591	7	ADB80145 Mycobacte
9	140	7.0	649	5	ABU05373 M. tuberc
10	139.5	7.0	367	7	ABO80543 Pseudomon
11	138.5	7.0	706	6	ABU22341 Protein e
12	136.5	6.9	372	7	ADM26375 Hyperther
13	133	6.7	944	6	ABR55194 Amino aci
14	132	6.6	3745	6	AAR37001 Micromono
15	130.5	6.6	3362	8	ADK16028 Streptomy
16	127.5	6.4	400	6	ABU20199 Protein e
17	126.5	6.4	851	3	AAB29627 Cat flea
18	126.5	6.4	1445	6	AAE32018 Human bio
19	126.5	6.4	3808	8	ADK16029 Streptomy
20	125	6.3	404	4	AA681206 Mycobacte
21	124.5	6.3	378	4	AAB96754 Putative
22	124	6.2	719	7	ABM87963 Rice abio
23	123	6.2	470	4	AAU48172 Propionib
24	123	6.2	470	6	ABM44691 Propionib

25	122	6.1	1029	8	AD15058 Pseudomon
26	122	6.1	1033	7	ABO71865 Pseudomon
27	120.5	6.1	457	4	AA690385 C glutami
28	120.5	6.1	472	9	ABM96727 M. xanthu
29	120.5	6.1	691	7	ABO70760 Pseudomon
30	120	6.0	508	7	ABO61024 Klebsiell
31	120	6.0	574	9	ABM91171 M. xanthu
32	120	6.0	886	9	ABM92782 M. xanthu
33	119.5	6.0	2969	2	AAW56446 Fragment
34	119	6.0	332	9	ABM96092 M. xanthu
35	119	6.0	413	6	ABU34130 Protein e
36	119	6.0	811	4	ABG29434 Novel hum
37	117.5	5.9	4635	8	ADK56098 Streptomy
38	117	5.9	707	8	ADR08714 Human pro
39	116.5	5.9	260	6	ABU33936 Protein e
40	116.5	5.9	517	2	AAW34204 Streptomy
41	116.5	5.9	517	2	AAW55805 Streptomy
42	116	5.8	810	9	ABM93215 M. xanthu
43	116	5.8	1402	4	AA681115 Mycobacte
44	116	5.8	1402	4	AAB66459 Protein e
45	115.5	5.8	386	8	ADN47126 Thermococ

## ALIGNMENTS

RESULT 1  
AD137307  
ID AD137307 standard; protein; 393 AA.  
XX  
XX AD137307;  
XX  
XX 22-APR-2004 (first entry)  
XX  
XX M. tuberculosis low oxygen induced antigen Rv2625c SEQ ID NO:28.  
XX  
XX mycobacterial infection; vaccine; tuberculosis;  
XX  
XX Mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;  
XX  
XX low oxygen induced antigen.  
XX  
XX Mycobacterium tuberculosis.  
XX  
XX WO2004006952-A2.  
XX  
XX 22-JAN-2004.  
XX  
XX 08-JUL-2003; 2003WO-DK000477.  
XX  
XX 13-JUL-2002; 2002DK-00001098.  
XX  
XX (STAT-) STRATENS SERUM INST.  
XX  
XX Andersen P, Rosenkrands I, Stryhn A;  
XX  
XX WPI; 2004-122778/12.  
XX  
XX N-PSDB; AD137352.  
XX  
XX Use of one or more polypeptides or their fragments, which are expressed  
XX during the latent stage of the mycobacterial infection, and/or nucleic  
XX acids encoding the polypeptides, for a therapeutic vaccine against  
XX tuberculosis.  
XX  
XX Claim 3; SEQ ID NO 28; 76pp; English.  
XX  
XX The present invention describes polypeptides or their fragments, which  
XX are expressed during the latent stage of a mycobacterial infection,  
XX and/or nucleic acids encoding the polypeptides, which are useful for  
XX creating a therapeutic vaccine against tuberculosis. Also described: (1)  
XX a therapeutic vaccine against tuberculosis comprising one or more  
XX polypeptides; (2) a method for treating an animal, including a human  
XX being, with tuberculosis caused by virulent mycobacteria, e.g. by  
XX Mycobacterium tuberculosis, M. africanum or M. bovis; (3) a method for  
XX immunising an animal, including a human being, against tuberculosis

CC caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis  
CC caused by virulent mycobacteria in an animal, including a human being;  
CC (5) a method for diagnosing previous or ongoing infection with a virulent  
CC mycobacterium; and (6) a method of diagnosing Mycobacterium tuberculosis  
CC infection in a subject. The polypeptides have antibacterial activities,  
CC and can be used in vaccines and in gene therapy. The polypeptides are  
CC useful for the manufacture of a therapeutic vaccine for treating an  
CC individual who is infected by a virulent mycobacterium, e.g. M.  
CC tuberculosis, and who is not vaccinated with BCG against tuberculosis.  
CC The present sequence represents a low oxygen induced antigen, which is  
CC used in the exemplification of the present invention.

XX Sequence 393 AA;

Query Match 100.0%; Score 1990; DB 8; Length 393;

Best Local Similarity 100.0%; Pred. No. 4,5e-182; Mismatches 0; Indels 0; Gaps 0;

Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRDAIPGRIAGFVNNHMSVLVILMLFTWSIATMLPGTVGGYPAVVYMLGAGAVML 60

DB 1 MRDAIPGRIAGFVNNHMSVLVILMLFTWSIATMLPGTVGGYPAVVYMLGAGAVML 60

QY 61 ASLAEHLAAHVAARAGVSVESTLWPGVTALGSAKTPKAFRIAPGPAISALS 120

DB 61 ASLAEHLAAHVAARAGVSVESTLWPGVTALGSAKTPKAFRIAPGPAISALS 120

QY 121 ATFGALITTLGAVRTPAIVISVAMWLATVNLGLFNLPGAPLDGRLVAYLRRHGD 180

DB 121 ATFGALITTLGAVRTPAIVISVAMWLATVNLGLFNLPGAPLDGRLVAYLRRHGD 180

QY 181 SVRAGIGARAGRVVALVIALGLAEFVAGGIVGVMIAFGWTFPAAREETRISTQQ 240

DB 181 SVRAGIGARAGRVVALVIALGLAEFVAGGIVGVMIAFGWTFPAAREETRISTQQ 240

QY 241 LFAGRVADAMTAOHTAPGWINVEDFQRYVLGSRHSAVPADDSITGLVALROLRD 300

DB 241 LFAGRVADAMTAOHTAPGWINVEDFQRYVLGSRHSAVPADDSITGLVALROLRD 300

QY 301 VAPSRSTTSVGDIALPLHSVPTARPOEPLTALERMAPLGPSPALVTEGSAAVGIYTP 360

DB 301 VAPSRSTTSVGDIALPLHSVPTARPOEPLTALERMAPLGPSPALVTEGSAAVGIYTP 360

QY 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

DB 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

QY 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

DB 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

QY 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

DB 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

QY 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

DB 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

QY 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

DB 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

QY 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

DB 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

QY 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

DB 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

QY 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

DB 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

QY 361 SDVARLIDVYRLAOPETFTTSPQADRRPSDAG 393

PI Huang LL, Larossa RA, McCluskey MP;  
XX WPI; 2002-619256/66.  
DR N-PSDB; ABS63259.  
XX

PS Claim 4; Page 49-50; 86pp; English.

XX The invention relates to a method for regulating gene expression of a

CC coding region in a Cyanobacterium comprising culturing the transformed

CC Cyanobacterium in the log phase or in the presence of UV-B light, where

CC the promoter region is activated and the coding region is expressed. The

CC method is useful for identification of highly expressed genes or UV

CC responsive genes, and their corresponding promoters. These genes and

CC promoters are useful for constructing expression vectors in

CC Cyanobacteria, green algae or plants and for the production of

CC biomaterials from sunlight. This sequence represents a high level

CC promoter polypeptide of the invention. (Updated on 29-AUG-2003 to

CC standardise OS field)

XX Sequence 379 AA;

QY 1 MRDAIPGRIAG--FVNNHMSVLVILMLFTWSIATML---PGTVGGYPAVVYMLGAGG 55

DB 11 MNNIRVGSLEGFIPFVNPVW--FLILGLVTLSTGODIARFPOLSGTP---MILGLIT 64

QY 56 AVMLASLAEHLAAHVAARAGVSVESTLWPGVTALGSAKTPKAFRIAPGPAISALS 115

DB 65 ALPLPASVVAHELHSLVALAOGIEVKSITLFLGSLASLEKESNTWQAFVAVIAPAV 124

QY 116 SLALSTATFGALATTLGAVRTPAIV--ISVAMWLATVNLGLFNLPGAPLDGRLVAY 173

DB 125 SLV-----FLGLITVGTQIPLPVPGQALIGLGMINLALALFLILGFLDGGNVLSKI 179

QY 174 LMRHGSVPRAGIGARAGRVVALVIALGLAEFVAGGIVGVMIAFGWTFPAAREE 233

DB 180 VMOITGQNNKXILIASRVGQFGWLAIALGSLGILNLPISGFWTILIGWFLONAGSSA 239

QY 234 TRISTQQLFAGRVADAMTAOHTAPGWINVEDFQRYVLGSRHSAVPADDSITGL 292

DB 240 RNAQVKQMEAFTRADVIFNSPIIPAGLNIREPANDVIGKTPWRFRFLVIGADNQLGV 299

QY 293 VALRLRDVAPSRSTTSVGDIALPLHSVPTAR--POEPLTALERMAPLGPSPALVTEG 351

DB 300 LATEDIK-----HVPTSDMPQVTVDSLMQY-----PQGMVTYNNAN 334

QY 352 SAVVGIVTPSDVARLIDVYRLAQ 374

DB 335 QSLF-----EVAQLDQOKLSE 351

QY 335 QSLF-----EVAQLDQOKLSE 351

DB 335 QSLF-----EVAQLDQOKLSE 351

QY 335 QSLF-----EVAQLDQOKLSE 351

DB 335 QSLF-----EVAQLDQOKLSE 351

QY 335 QSLF-----EVAQLDQOKLSE 351

DB 335 QSLF-----EVAQLDQOKLSE 351

PN US6833447-B1.  
XX  
XX 21-DEC-2004.  
XX  
XX 10-JUL-2001; 2001US-00902540.  
XX  
XX 10-JUL-2000; 2000US-0217883P.  
XX  
XX (MONS ) MONSANTO TECHNOLOGY LLC.  
XX  
XX Goldman BS, Hinkle GJ, Slater SC, Wiegand RC,  
XX WPI; 2005-028716/03.  
XX  
XX New substantially purified *Myxococcus xanthus* nucleic acid molecule  
XX encoding a nitrite reductase, useful for determining gene expression,  
XX identifying mutations in a gene of interest, and for constructing  
XX mutations in a gene of interest.  
XX  
XX Example 2; SEQ ID NO 16040; 25pp; English.  
XX  
XX The invention relates to a substantially purified nucleic acid molecule  
XX encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a  
XX recombinant DNA construct for expression of a nitrite reductase gene in a  
XX plant cell, and a plant cell comprising the recombinant DNA construct.  
XX The nucleic acid is useful for determining gene expression, identifying  
XX mutations in a gene of interest, and for constructing mutations in a gene  
XX of interest. Sequences given in records for SEQ IDs 9692-16825 represent  
XX a group of 7134 *Myxococcus xanthus* proteins and peptides. Note: The  
XX sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from USPRO  
XX  
XX Sequence 290 AA;  
SQ

Query Match 12.3%; Score 244; DB 9; Length 290;  
Best Local Similarity 30.5%; Pred. No. 1.3e-14;  
Matches 91; Conservative 48; Mismatches 117; Indels 42; Gaps 11;  
QY 88 LFGGVTALGSEAKTAKAFRIAPAGPATSLATGALATLACVTRPAIVISAMWLA 147  
DB 2 MGVGVSELTETAPRRRDEALMAAVGLTSLILAVLIGATWLLQEVRSNLOFAC-PTMA 60  
QY 148 TVNLLGLFNLPGAPLDGRLVRAYYLWRHGDVVRAGIGAPAGRVVALVILALGAEF 207  
DB 61 SLNLFGLGFNLPPAPMDGRLVRASLAGRLG-MVRAIQVASWLRGFAVL- 111  
QY 208 VAGGLVGW- - - - -LAFIGMTFPAAREESTRISTQULFAGRVADANTAOPTAP 259  
DB 112 - - - - -GVMAVLSLNPFLLVIAFPIFMGAGCEAOQVMKTLERVPVADLMTTPRRVGD 164  
QY 260 GMINVEDFI- - - - -QRVYLSERHSAYPVADRGSITGLVALRQLRDVAPSRSTTSVGD 314  
DB 165 AGASIEQALMDLRRLRL- - - - -LPYTE-DERPVGVSLETVAVAPDSEMRITVEV 217  
QY 315 ALPLHSVPTAAPQEBFTALRLERMAPLGRSPALVTESAVV-GIYTPSDVALIDVVR 371  
DB 218 MVP- - - - -AVVVRIDEDQWTA-LRRMAE-EERPQLVVLEADGVLAGTIDVADVGKMALYQ 271

RESULT 4  
ABM94210  
ID ABM94210 standard; protein; 346 AA.  
XX  
XX ABM94210;  
XX  
XX 02-JUN-2005 (first entry)  
XX  
XX M. xanthus protein sequence, seq id 13409.  
XX  
XX Transgenic plant; DNA replication; gene regulation; gene expression.  
XX  
XX *Myxococcus xanthus*.  
XX

PN US6833447-B1.  
XX  
XX 21-DEC-2004.  
XX  
XX 10-JUL-2001; 2001US-00902540.  
XX  
XX 10-JUL-2000; 2000US-0217883P.  
XX  
XX (MONS ) MONSANTO TECHNOLOGY LLC.  
XX  
XX Goldman BS, Hinkle GJ, Slater SC, Wiegand RC,  
XX WPI; 2005-028716/03.  
XX  
XX New substantially purified *Myxococcus xanthus* nucleic acid molecule  
XX encoding a nitrite reductase, useful for determining gene expression,  
XX identifying mutations in a gene of interest, and for constructing  
XX mutations in a gene of interest.  
XX  
XX Example 2; SEQ ID NO 13409; 25pp; English.  
XX  
XX The invention relates to a substantially purified nucleic acid molecule  
XX encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a  
XX recombinant DNA construct for expression of a nitrite reductase gene in a  
XX plant cell, and a plant cell comprising the recombinant DNA construct.  
XX The nucleic acid is useful for determining gene expression, identifying  
XX mutations in a gene of interest, and for constructing mutations in a gene  
XX of interest. Sequences given in records for SEQ IDs 9692-16825 represent  
XX a group of 7134 *Myxococcus xanthus* proteins and peptides. Note: The  
XX sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from USPRO  
XX  
XX Sequence 346 AA;  
SQ

Query Match 9.7%; Score 174; DB 9; Length 346;  
Best Local Similarity 27.4%; Pred. No. 8.6e-08;  
Matches 92; Conservative 37; Mismatches 137; Indels 70; Gaps 17;  
QY 58 MLASLALHELAHVVARBAVSVESTLTMFGVTALGSEAK-TPKAAPRIAPAGPAT 116  
DB 19 IVFVSVLHGLHGLALAPRRYCCPAR-TELHMGSTTQHDNAHLTHQOSATVSRGPGIG 77  
QY 117 LALSATPGALA- - - - -ITLAGVTRPAIVISVAMWLTATVNLGLFNLPGAPIDG 166  
DB 78 FLAGGLTWGLSGLVPLBPGGLADQV- - - - -QFLM- - - - -VNIGMGLFVLLPMDG 127  
QY 167 GR- - - - -LYRAYLWRHGDVVRAGIGAPAGRVVALVILALGAEFVAGLVGVTWLAFIG 222  
DB 128 GHLADLVRAASGYRHERGV-LGVG- - - - -ITAVVVLGLAIW-SKOLMMGMLAWVLG 178  
QY 223 WFIFAAREESTRISTQULFAGRVADANTAOPTAP-GWINVEDFIQRVYLSERHSAYP 281  
DB 179 VMNIEQLRRTKRRPAERFALPR- - - - -LVKRPDAAPRGALISIQLDL- - - - -RGTRRP 230  
QY 282 - - - - -VADRGSITG- - - - -LVALRQLRDVAPSRSTTSVGDIALPLHSVPTAAPQEB 330  
DB 231 PGADADDDLEGRPDPLVGEMLDNGLE- - - - -LAVGSLQSAFTQAPLARTGALVIAL 286  
QY 331 - - - - -TALLERMAPLGRSR- - - - -ALVTESAVV 355  
DB 287 LHTGRLOGLASILDSSHARQLSEDTLALISRHACTV 322

RESULT 5  
ADN48111  
ID ADN48111 standard; protein; 438 AA.  
XX  
XX ADN48111;  
XX  
XX 01-JUL-2004 (first entry)  
XX  
XX *Thermococcus kodakaraensis* KOD1 protein sequence SeqID1989.  
XX  
XX



KM gene disruption; gene targeting; marker gene; transformation;  
 KM homologous recombination; hyperthermostable archaeobacterium; KOD1;  
 KM gene structure; gene function; enzyme activity; medicine;  
 KM forensic science; food; drug inspection; molecular biology; immunology.  
 XX  
 OS Thermococcus kodakarensis.  
 PN WO2004022376-A1.  
 PD 18-MAR-2004.  
 XX  
 PF 29-AUG-2003; 2003WO-1B003597.  
 XX  
 PR 30-AUG-2002; 2002JP-00319011.  
 XX  
 PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
 XX  
 PI Imanaka T, Atomi H;  
 DR WPI; 2004-257583/24.  
 PT Method for disrupting targeted gene in genome of organism particularly  
 PT thermostable bacterium and with genome chips for analysis, applicable in  
 PT studying gene structure and functions.

PS Claim 9, SEQ ID NO 1989; 598pp; Japanese.  
 XX  
 CC This invention relates to a novel method for targeting disruption of an  
 CC arbitrary gene in a genome of an organism which comprises providing the  
 CC whole sequential data of the genome of such organism, selecting at least  
 CC 1 arbitrary region in the sequence, providing a vector that contains a  
 CC sequence homologous with the selected region and a marker gene,  
 CC transformation, and homologous recombination. The genome is preferably  
 CC the genome of a hyperthermostable archaeobacterium, particularly  
 CC *Thermococcus kodakarensis* KOD1. The method is for targeting the  
 CC disruption of a gene in the genome of an organism, which is applicable in  
 CC studying gene structure and functions as well as enzyme activities of  
 CC encoded proteins and useful in medicine, forensic science, food or drug  
 CC inspection, molecular biology and immunology. With this method, the  
 CC disruption of a gene at an arbitrary position in a genome can be achieved  
 CC efficiently and reliably. The present sequence is that of a protein  
 CC encoded by the genome of *Thermococcus kodakarensis* which was derived  
 CC using the method of the invention. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 CC Sequence 438 AA;

[illegible]

```

QY      244 GVR--VADAMTAQPHHTAPGW 261
          : | : | | |
Db      419 ALTALVIFVITATPR--PLW 436

```

RESULT 6  
AAB96393  
ID AAB96393 standard; protein; 409 AA

AC AAB96393;

DT 29-OCT-2001 (first entry)

Putative P. abyssi membrane Zn protease #1.

KW Hyperthermophilic archaeon; hyperthermophilic protein

OS *Pyrococcus abyssi*.

PN FR2792651-A1  
 YY

PD 27-OCT-2000.  
XX

PF 21-APR-1999; 99FK-00005034  
XX XX

PR	21-AFK-1999;	99FK-00003034
XX		

PA (IFRE-) IFREMER INST FR RECH EXPL MER.

PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O,

XX

XX

PT useful in industry.

PS Claim 7; Page 1083-1084; 1657pp; French

CC The present invention relates to the genomic sequence of *Pyrococcus*  
CC *abyssi* (see A4966431 and A441423-7) and P. *abyssi* proteins. P. *abyssi* is  
CC a hyperthermophilic archaeon, which is isolated from deep-sea  
CC hydrothermal vents. The present invention has one such P. *abyssi* protein,  
CC the proteins of the present invention have various potential industrial  
CC uses, since the proteins are stable at very high temperatures, some up to  
CC 110 degrees centigrade. Note: This patent is in the same patent family as  
CC WO2000065062, which contains additional sequences as shown in A4939132-  
CC A4939143, A4475903-A4475920 and A4665436

**SQ Sequence 409 AA;**

Query Match 7.2%; Score 143; DB 4; Length 409;  
Best Local Similarity 20.7%; Pred. No. 0.0001;  
Matches 59; Conservative 47; Mismatches 83; Indels 96; Gaps 10.

11 AGFVNTHMSVLVILMLFTMSLATM-LPGTVGGYPVVYWL LGAGGAVMLLASLLAHEL

Db 142 AGYILST-----LYVTTLLEENLPGIKNTYLNALAFSLG-----IISILGTHEM

QY 70 HAVVARAGVSVEVT-----LWLEGGVTALGGEAKTPKAFRIAFAGPATS

Db 187 HKIAA-----SIHNVKSTFPYFI PPSPFICTLGAVIRVKSPIPTRNAEVDLGVSGPIAG

QY 118 ALSATFGALATLACVTRPAIV-----

Db 242 LV-----AIPVTIIGLKSAVVPINYLEKETIYFGSSLLFYGLMKLVLGDLPPQNVGII

```

140 --ISVAMMLATVNLGLFNLPGAPLDGGRLVRLWRRHGSVRAGIGARAGRVA

```

DB 29 / HPLAVAGM---VGLVIFLNLLEAQLDDGSHVARALL-----PENAKKVL

136 VL--LAWGHEF VAGGUVGVWUAF IGMF IFMMWKEEINIOIQ 270

Db 341 TLGFLTIGLAFWPGMILMGILILMGRVGNFGALDEVSPILTTSR 365

RESULT 7

ID ADM26447 standard; protein; 177 AA.

AC ADM26447;

DT 20-MAY-2004 (first entry)

DE Hyperthermophile Methanopyrus kandleri protein #1053.

KW hyperthermophile; protein stability enhancement;

KM protein activity enhancement.

OS Methanopyrus kandleri.

PN W02003076575-A2.

PD 18-SEP-2003.

PF 04-MAR-2003; 2003WO-US006664.

PR 04-MAR-2002; 2002US-0361742P.

PR 14-MAY-2002; 2002US-0380423P.

PR 16-SEP-2002; 2002US-0410974P.

PA (FIDE-) FIDELITY SYSTEMS INC.

PA (MALY/) MALYKH A.

PI Slesarev AI, Pavlov A, Pavlova N, Kozyavkin S;

DR WPI; 2003-748383/70.

DR N-PSDB; ADM27081.

PT New isolated nucleic acids encoding any of about 1700 Methanopyrus

PT kandleri proteins, and the encoded proteins, useful as a medicaments or

PT as diagnostic agents.

PS Claim 31; SEQ ID NO 1053; 1023bp; English.

XX The invention comprises the amino acid sequence of proteins from the

XX hyperthermophile Methanopyrus kandleri, the invention also comprises the

XX complete genome from Methanopyrus kandleri. The Methanopyrus kandleri

XX proteins of the invention are useful for enhancing the stability and/or

XX activity of other proteins. The Methanopyrus kandleri genome is useful in

XX a variety of diagnostic and analytical methods. The present amino acid

XX sequence represents a Methanopyrus kandleri protein of the invention.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX ADB80145;

DT 04-DEC-2003 (first entry)

DE Mycobacterium tuberculosis nutrient starvation-inducible protein #92.

KW mycobacteria; nutrient starving condition; mycobacterial latency;

KM mycobacterial infection; vaccine.

OS Mycobacterium tuberculosis.

PN W02003004520-A2.

PD 16-JAN-2003.

PF 04-JUL-2002; 2002WO-GB003052.

PR 04-JUL-2001; 2001GB-00016385.

PR 05-OCT-2001; 2001GB-00023993.

PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.

PA James BM, Marsh P, Hampshire T;

DR WPI; 2003-210338/20.

DR N-PSDB; ADB80146.

PT New mycobacterial peptide, useful for the manufacture of a medicament for

PT treating or preventing, or a diagnostic reagent for identifying,

PT mycobacterial infection.

PS Claim 2; Page 312-315; 442bp; English.

XX The invention comprises the amino acid and coding sequences of

XX mycobacterial (Mycobacterium tuberculosis) proteins which are up-

XX regulated under nutrient starving conditions and maintain mycobacterial

XX latency. The DNA and protein sequences of the invention are useful for

XX the treatment, prevention and diagnosis of a mycobacterial infection. The

XX present amino acid sequence represents a Mycobacterium tuberculosis

XX protein of the invention.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Db 375 ASPARAATVASAVLFTSSLYSTATFLASWDFNTEBYLKNQASLAAASGAPLLDQE 434  
Qy 369 -----VYRLAQPE-----PTFTTSPQDADRFSDAG 393  
Db 435 VDPVLQGVAMPENLASHMFALLRVRPEFATTTQLRMFSTGTG 477

RESULT 9  
ABU05373  
ID ABU05373 standard; protein; 649 AA.  
XX  
AC ABU05373;  
XX  
DT 08-APR-2003 (first entry)  
XX  
DE M. tuberculosis and M. leprae marker protein #24.  
XX  
KM Mycobacteriosis; survival; virulence; protective antigen; vaccine;  
KM Mycobacterial disease; tuberculosis; leprosy.  
XX  
OS Mycobacterium tuberculosis.  
OS Mycobacterium leprae.  
XX  
FN WC200274903-A2.  
XX  
PD 26-SEP-2002.  
XX  
PF 22-FEB-2002; 2002WC-IB001973.  
XX  
PR 22-FEB-2001; 2001US-0270123P.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
PI Cole S;  
XX  
XX WPI; 2002-755985/82.  
XX  
PT Identifying and selecting genes for survival or virulence of mycobacteria  
PT by a comparative genomic analysis of the sequences of Mycobacterium  
PT tuberculosis and M. leprae.  
XX  
PS Claim 17, Page 190-192; 874pp; English.  
XX  
XX This invention relates to a novel method for identifying essential genes  
CC for survival or virulence of mycobacteria species. The method comprises  
CC aligning the genomic sequence of a first mycobacterium species on a  
CC genomic sequence of a second mycobacterium species and selecting a  
CC polynucleotide sequence that is highly conserved in both genomes with no  
CC counterparts in other bacterial genomic sequences and that corresponds to  
CC an essential gene for the survival or virulence of mycobacterium species.  
CC The method of the invention is useful for detecting M. tuberculosis or M.  
CC leprae infection. The method reduces the number of potential new targets  
CC and protective antigens for new drugs and vaccine compositions to treat  
CC and prevent mycobacterial diseases, particularly tuberculosis and  
CC leprosy. The present sequence represents a marker protein from  
CC Mycobacterium tuberculosis and Mycobacterium leprae identified using the  
CC method of the invention  
XX  
SQ Sequence 649 AA;  
Query Match 7.0%; Score 140; DB 5; Length 649;  
Best Local Similarity 23.5%; Pred.No. 0.00037;  
Matches 114; Conservative 62; Mismatches 172; Indels 138; Gaps 25;

Qy 11 AGFVNNHMSVLYVIM--LFTWSL-----ATMLP-----GTVG 41  
Db 105 AAVLVSVGMVLMIGTGRLYAPSLQGRQATMLPAGALVVALPPARDPATSGESGLVL 164  
Qy 42 GYPAVVVWL-----LGAGGAVMLASLAEHLAAHVAVARRAGVSVEV 84  
Db 165 TYLGILMMVMVCMQAPLRNRSQSRRTGALAFVAGCSVLVPEELAMGSGSLIMLMIAAR 224

Qy 85 TLMFGVTLGGBAKTPKAFRIAFAG---PATSLALST---FGMLATTLGAVRTPAI 138  
Db 225 TCWLRALIVVAGGSLPVAYOUFRMGYGLLVPGTALAKDAADKMSGCIITYLSNFQPYV 284  
Qy 139 VISVAMWLATVNL-LGLFNL-----PGAPLDGGR-----L 169  
Db 285 L-----WVPLVLVLGLLMLLHRMPSFMHPLETPESSGVAARAVOSPAAVVVFFVPSGL 339  
Qy 170 VRAYLMRR-----HGDSVRAGIGAAARAGVVALVLIALLGLAEF--VAGGLVGV---W 217  
Db 340 LQAFYVIRGGDFHMGHGVLLAPFLCLLAPVAVI PVVISEG-ADFSRQTGMWLAGVSLTW 398  
Qy 218 LAFIGWFIAPA-----AEEETRISTOQLFAGVAVADMTAP--HTAPGININDEFIQRY 271  
Db 399 LGVAGMSLMAANSPGMDATNVS---YSGIDERRFFVQATGHAP--LTAADYV--- 449  
Qy 272 VLGERHSAVPAVADGSIITGLVALRQUDVAPRRSTSTVGDIAPHSVPTARPOEPLT 331  
Db 450 -----GYF---RMAAV--LVALLNTPDQALLPSGNTYIKMDLVPMIQSPSPSPSPD 497  
Qy 332 ALLERMAPLQPRSRALVTGSAVVGIVTPSDVARLIDVYRLAQPEPTTSPQ-----D 385  
Db 498 SLVSGK----PQHTVFPT-NIGMLGMNVGADV-RVIDQIGLANPLAQHTERLQHGRIQHD 551  
Qy 386 ADRFSD 391  
Db 552 KNLEPP 557

RESULT 10  
ID AB080543 standard; protein; 367 AA.  
XX  
AC AB080543;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Pseudomonas aeruginosa polypeptide #12718.  
XX  
KM Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.  
XX  
OS Pseudomonas aeruginosa.  
XX  
FN US6551795-B1.  
XX  
PD 22-APR-2003.  
XX  
PF 18-FEB-1999; 99US-00252991.  
XX  
PR 18-FEB-1998; 98US-0074788P.  
XX  
PR 27-JUL-1998; 98US-0094190P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
XX WPI; 2003-615309/58.  
XX  
DR N-PSDB; ABD14114.  
XX  
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 29289; 455pp; English.  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant

CC production of *P. aeruginosa*-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of *P. aeruginosa*-caused  
CC infection, and in detection of *P. aeruginosa* sequences or other sequences  
CC of *Pseudomonas* species using biochip technology. Sequences AB067826-  
CC AB084336 represent *P. aeruginosa* polypeptides of the invention. Note: The  
CC sequence data for this patent did not form part of the printed at  
CC specification but was obtained in electronic format from USPTO at  
CC [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html)

XX Sequence 367 AA;

Query Match 7.0%; Score 139.5; DB 7; Length 367;  
Best Local Similarity 24.9%; Pred. No. 0.00019;  
Matches 98; Conservative 48; Mismatches 143; Indels 105; Gaps 18;

OY 24 ILMLPTWMLATMLPTGVGYPAVVY-----MLLAGAGVMLLASLHSLHAAVVARAG 78  
DB 27 VTKRLYV--ATPLVAGALGMLASLVGWPLPMWIGSLAVMLVRCIADLPFAEVPARKKG 84  
OY 79 VSVESVTLMFG-GVTALGGEAKTPKAAFRFAAGPATSLASA--TFGALATITAGVRT 135  
DB 85 -----QWVGIGI-----GLHFTPAVIEQV-----LAHSVITIVGAVATITLSSVLA 125  
OY 136 PAIVTSVAMWMLATVNLGLFNLPGAPLDGRLVRAVILMRKGDVYAGAGAAAGV 195  
DB 126 IAFMRSGEDRATAPF-----ASMPGASBMVNLGGRHG-AVLSRVAAAGSLRL 174  
OY 196 ALVIALGLAEFVAGLVGV-----MLAF-----IGMFFFAAREBEET-- 235  
DB 175 LVVLVLPVPAFGTLGGGGRPHQAAPVDWMLLALFPAGALVALDW-----OKLRQP 226  
OY 236 ----ISTOOLPAGVVA--DAMTAOPHTAG--WINEVDPIORVYLGRHSAYPVADRQ 287  
DB 227 NPWLLGPLLAAVGLGFDLHGLPAGSSGVQWL-----IG 263  
OY 288 SITGVALRQLRDVAPSRRTTSVGDIALPLHSVPTAPPOEPLT-----ALIERAPLGP 342  
DB 264 SALGCHFNRSFPRSPAPFVSRITVCTLMMFMAALAAELGLWTTLTHQSIMLGMPGSI 323  
OY 343 RSRALVTGSAY-VGIVTPSDVAPRLIDVYRLAQP 375  
DB 324 AELSLTAEALQISVPLVLTALQVLRLLVFLAEP 357

RESULT 11

ABU22341  
ID ABU22341 standard; protein; 706 AA.

XX AC ABU22341;

DT 19-JUN-2003 (first entry)

DE Protein encoded by prokaryotic essential gene #7868.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Burkholderia mallei.

XX WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002MO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Hasebeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-023926/02.  
DR N-PSDB; ACA26211.

XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 50265; 1766bp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibody; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
XX [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)

XX Sequence 706 AA;

Query Match 7.0%; Score 138.5; DB 6; Length 706;  
Best Local Similarity 22.2%; Pred. No. 0.00057;  
Matches 110; Conservative 48; Mismatches 159; Indels 179; Gaps 21;

OY 7 LGRIAGFVVNV---HWSVLVTLMLFTWMLPTGVGYPAVVYMLGAGVMLASL 63  
DB 121 LGLTAAMLVPLAAGHASY-PLTFVCGLAACVSAAGGMRATPLRLILAGSVCMLLFSA 179  
OY 64 LABEL-----AAHVARRAGVSVESVTLMFGVTLG-----GE 98  
DB 180 VTTLLAFFEQTVVAGSLMWSGSLYOPGAGLRDALMLVAPLALVLRPLDPLAGD 239  
OY 99 AKTPKAAFRFAAGPATSLA---LSATFGALATITAGVTPAIVTSAMWMLATVNL-- 152  
DB 240 DAAAAAGVRV---DATRLAGTVVAVGFAVAVSIAG--PLSYTG---LIAPLIRQM 288  
OY 153 -----LGLFNLPGAPLDGRLV----- 170  
DB 289 RGAKASRLGA--LVPLAALAGALVVTDSAVLALGIDATISTGVAATAPVOTPLMLAMIR 346  
OY 171 PAVLYW---RRHGDVSRAGI--GAAPAGR-----VALVITIALGLAEFVAGLVGVWL 218  
DB 347 RGAAMSGAQQPAHVRAAGTRGAGALVRLAVLPMPAVAVLILAAAGAVLVVVGSAFGPWI 406  
OY 219 AFIGMFFFAAREBEET-----ISTQOLF----- 242  
DB 407 GPRGFPAALQRDRLARTVLDLRAPRLCALLAGALLGASGVLMQSTVRNPLAGPVLGV 466



Db 529 RAIPPKKIVTV-LVGGTALALDISHGLFAKMP--LMVVIILTTTIVIMLAFGSVLP 584  
Qy 130 LAGVTPAIVISVAMMLATVNLGLF-----NLIP---GAPLDG----- 166  
Db 585 IKATLMSALTLG-----STMGLITWIFVDGHSKMLNTPPLTKPVLGIIIALVFGIST 639  
Qy 167 -----GRLVAYIMRRHGDVSRAG--IGAARAGRV--ALVIALGLAEFVAGLVG 215  
Db 640 DVEAVELVRMVEA---RRGMSSTQGAIRIGTAQGRITTAALIVAVVGAFFVSDLVMM 696  
Qy 216 VWLAFIGFIFRAAAEEETRISTOOLFAGVRYADAMTQPHTRPW-----IN 263  
Db 697 KYLA---GLMAALLDLATV---RMFLVPSVMKLLGDCCWMAFRMARLQTRIGLGIH 750  
Qy 264 VEDFIQRYVYGERHSAVPAVDRDGSITGLVALRQLRDVAFSRSTVSVDIALPLHSVPT 323  
Db 751 LPDEKRVSNRPARPPVT-----AGLVAKARAAGDRPPHDPHPLAESRPARSSPA 804  
Qy 324 ARPOEPLTALLERMAPLGPSPALVTESAVVGIPTSDVABLIDVRLAQ-----PPT 378  
Db 805 SSPE--LTPALEATAPAPASGASTTR--MQIGSSTPTPTRLAAAGRSVQSPASTPPT 860  
Qy 379 FT-----TSPQADRFSDA 392  
Db 861 PTPSPAPSAQOTRAMPLAANRSTDA 885

RESULT 14  
AAE37001  
ID AAE37001 standard; protein: 3745 AA.  
AC AAE37001;  
XX  
DT 23-OCT-2003 (revised)  
DT 07-AUG-2003 (first entry)  
XX  
DE Micromonospora carbonacea polyketide synthase (PKS) type I #7.  
XX  
KW Macrolide; rosamycin; polyketide; polyketide synthase; PKS; enzyme.  
XX  
OS Micromonospora carbonacea.  
XX  
PN CA2391131-A1.  
XX  
PD 19-NOV-2002.  
XX  
PF 26-JUN-2002; 2002CA-02391131.  
XX  
PR 26-JUN-2001; 2001US-0307629P.  
XX  
PA (ECOP-) ECOP1A BIOSCIENCES INC.  
XX  
PI Yang X, Staffa A, Farnet CM;  
XX  
DR WPI; 2003-343556/33.  
DR N-PSDB; AAD55817.  
XX  
PT Novel isolated polypeptide involved in biosynthesis of macrolides by  
PT microorganism, useful for biosynthesis of macrolides by microorganisms,  
PT preferably for biosynthesis of rosamycin.  
XX  
PS Claim 13; Page 138-151; 206pp; English.  
XX  
CC The invention relates to genes and proteins involved in the biosynthesis  
CC of macrolides by microorganisms. In particular it relates to the nucleic  
CC acids forming the biosynthetic locus for rosamycin (a 16-member  
CC macrolide antibiotic) from Micromonospora carbonacea. The invention is  
CC useful for the biosynthesis of macrolides by microorganisms. It allows  
CC direct manipulation of macrolides and related chemical structures by  
CC chemical engineering of the proteins involved in the biosynthesis of  
CC rosamycin. It is useful to catalyse certain biochemical reactions, in  
CC vitro or in vivo, to direct or enhance the synthesis or modification of a  
CC polyketide, polyketide substrate or its precursor. The present sequence

CC is M. carbonacea polyketide synthase (PKS) type I protein. (Updated on 23  
CC -OCT-2003 to standardise OS field)  
XX  
SQ Sequence 3745 AA;  
Query Match 6.6%; Score 132; DB 6; Length 3745;  
Best Local Similarity 25.2%; Pred. No. 0.022;  
Matches 107; Conservative 40; Mismatches 174; Indels 104; Gaps 20;  
Qy 31 SLATMLPGTGVGVAVVYVWMLGAGVWMLASLT-AHELIHAVVARRAGVSES----- 83  
Db 1063 SLMLAEAGADRRHRAVP---LGLTASLALIQALADAGTQARLMVYTRGAVAVSSGEVPDA 1119  
Qy 84 --VTLMFGVYVTAAL-----GGEAKTPKAFRIAFAGPATSLASATFGALATLTAAG--- 132  
Db 1120 GQAGVWGLGRVVALELDPKMGVLVDLPALNGERAFQALADVVGSGNGEDQAVAVASGVYG 1179  
Qy 133 ---VTPPAIVISVAMMLATVNLGLFNLFGAPLDGRLVRAVLMRRHGDVSRAGICAA 189  
Db 1180 RRLVRSRATVYSGDMPARGTILVVG--DTGVAALLAGRLV-----GDGAA 1223  
Qy 190 R---AGRVVALVIALGLAEFVA-----GGLVG-----GWLAFIG 222  
Db 1224 HVTLAEPAASTVGLTGADRVVALIDCDPSRDALAGLGAYPRTTVVAPPAVALTLA 1283  
Qy 223 W-----FIFAAREEETRISTOOLFAGVRV-ADAMTQPHTRPQMINVEFIQRYVGER 276  
Db 1284 ETPPEDFVAIVAAKTTTVAHDLAAEALELDAFVVFSSVSGTW-----GGAG 1332  
Qy 277 HSAYPVADRDSITGLVALRQLRDV-----APSRSTSVG---DIALPLHSVPTAR 325  
Db 1333 HGCV--AAGTARLALVEERRARGLPATJAMTWMADATTAAAGCAPDASAGHE-PPTR 1389  
Qy 326 PQEPLTALLER--MAPGPRSRALVTESAN--VGIYTPSDV--ARLIDVYRLAQPEPT 379  
Db 1390 AGPDRBLRRGGILPLDPGALDVLKGAVARGEGLTVADVDMARFVASTAARPTTLF 1449  
Qy 380 TTSPO 384  
Db 1450 DELPE 1454

RESULT 15  
ADK16028  
ID ADK16028 standard; protein: 3362 AA.  
XX  
AC ADK16028;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE Streptomyces halstedii vincenistatin VinP3 product seq id 6.  
XX  
KW cytosolic; gene therapy; gene cluster; antibiotic; vincenistatin;  
KW dUDP-vicenisamine; cancer; VinP3; polyketide synthase.  
XX  
OS Streptomyces halstedii.  
XX  
PN US2004053274-A1.  
XX  
PD 18-MAR-2004.  
XX  
PF 04-MAR-2003; 2003US-00378083.  
XX  
PR 04-SEP-2002; 2002JP-00258683.  
XX  
PA (TOKD ) TOKYO INST TECHNOLOGY.  
XX  
PI Kakinuma K;  
XX  
DR WPI; 2004-247723/23.  
DR N-PSDB; ADK16023.  
XX  
PT New gene clusters encoding a vicenisamine glycosyltransferase polypeptide

PT useful in the biosynthesis of vicenistatin using dTDP-vicenosamine as a  
PT substrate and for treating cancer.

PS Claim 8; SEQ ID NO 6, 103pp; English.

XX  
CC The invention describes a gene cluster comprising a base sequence  
CC represented by base numbers 1-64492 of a 64492 base pair sequence (SEQ ID  
CC No. 1) encoding proteins required in the production of antibiotic  
CC vicenistatin, given in the specification. Also described are the  
CC following: a polypeptide comprising a 419 amino acid sequence (SEQ ID No.  
CC 4), given in the specification, or an amino acid sequence having at least  
CC 90 % homology with SEQ ID NO: 4, and exhibiting enzymatic activity as a  
CC glycosyltransferase; a gene encoding a polypeptide of (1); a gene  
CC comprising a base sequence represented by base numbers 2790-4049 in SEQ  
CC ID No. 1, a base sequence having at least 90 % homology with the base  
CC sequence cited above, and encoding a polypeptide exhibiting enzymatic  
CC activity as a glycosyltransferase, or a gene comprising a base sequence  
CC that hybridizes with the base sequence cited above under a stringent  
CC condition, and encoding a polypeptide exhibiting enzymatic activity as a  
CC glycosyltransferase; a polypeptide comprising an amino acid sequence  
CC represented by a 355, 323, 2260, 3362, 3808, 236, 397, 493, 330, 158,  
CC 469, 299, 327, 82, 524, 478, 414 or 5835 amino acids (SEQ ID NO: 2-3 or 5  
CC -20), respectively, and producing vicenistatin by catalytic reaction of  
CC the glycosyltransferase, comprising the polypeptide of (1), using dTDP-  
CC vicenosamine as a starting material. The methods and compositions of the  
CC present invention are useful for producing vicenistatin in the treatment  
CC of cancer. This is the amino acid sequence of a polypeptide synthase  
CC encoded by the Vinp3 gene of the Streptomyces halstedii gene cluster  
CC comprising genes encoding proteins required in the production of the  
CC antibiotic vicenistatin.

XX Sequence 3362 AA;

Query Match 6.6%; Score 130.5; DB 8; Length 3362;

Best Local Similarity 24.1%; Pred. No. 0.027; Mismatches 153; Indels 69; Gaps 14;

Db 47 VYMLGAGGAVVMLASLAEHLAAVVARAGV-----VESVTLMFGVTA 94  
416 VDM---SAGAVELTLE--AREWPTGRPRRAGVSFGSGTNAHYIEASEPEPSAVER 470  
QY 95 LGGEAKTP-----KAAFRIFAGPAPATSLASATFGALATTLGAVTPRAVI 140  
Db 471 LAGGVTTPPWVLSARSADALRGQERLLSFVSAAGDVSVDAVYSLGVSRAGLEHRGVV 530  
QY 141 SVAMMLATVNLGLFNLPGAPLDG-----GRLVRAVYLMRRHGDSPVAGIGARAG 192  
Db 531 GE---SRAELLALESLSAGVESPGVVTGRVABGR--AFVFTGCG--AQRVGMRELA- 582  
QY 193 RVVALVLIALGLAEFVAGLVGVWLAFIWMFIPAAAREESTRISTOQLFAGVRVADAMT 252  
Db 583 --AAFPFPAASLEE--TCGLLEBAGVA--VREVLFAEBSGAELALLTRVYAQAALFAVEV 637  
QY 253 AQHTAPGWINVEDFIQRYVIGERHSAY--PYADRDGSITGLVALROLDAVPSRRSTTS 310  
Db 638 ALFRLVESFGVVPDVFAGHSVGEIAAAHVAGFSLIEDAVSLVAARGRLMDALPEGGAIVA 697  
QY 311 VGDIALPLHSAVPTARPOEPLTALLERMAPLGRSRALVTESGSAVVGIYTPSDVARLIDVY 370  
Db 698 V-----QATEBDVALLEGEV---DASIAIINGPDVAVVSGTEAGVARVVDVL 742  
QY 371 R 371  
Db 743 R 743

Search completed: March 23, 2006, 05:10:54  
Job time : 107.486 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 05:11:18 ; Search time 15.6356 Seconds  
(without alignments)  
2418.401 Million cell updates/sec

Title: US-10-617-038-28

Perfect score: 1990  
Sequence: 1 MRD1PILGR1AGFVVNVHMS.....QPEPTFTTSPODADRFS DAG 393

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Fred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1990	100.0	393	2	H70572
2	486.5	24.4	389	2	AP2036
3	477	24.0	399	2	AE2293
4	459	23.1	390	2	D84367
5	425.5	21.4	379	2	S76029
6	418	21.0	362	2	D69291
7	377.5	19.0	341	2	A69209
8	328.5	16.5	339	2	H64348
9	328	16.5	372	2	H95972
10	288	14.5	372	2	A95338
11	173	8.7	365	2	C64421
12	149.5	7.5	284	2	E97054
13	144	7.2	493	2	AD2070
14	143.5	7.2	373	2	D71142
15	143	7.2	409	2	B75010
16	143	7.2	503	2	S77290
17	142.5	7.2	2569	2	T14164
18	140	7.0	591	2	D70562
19	140	7.0	649	2	H66920
20	139.5	7.0	345	2	D83551
21	139	7.0	191	2	C64376
22	136.5	6.9	289	2	G84026
23	134	6.7	2129	2	T14182
24	133.5	6.7	199	2	A69232
25	133	6.7	944	2	C70839
26	132.5	6.7	360	2	C82445
27	130.5	6.6	871	2	AC2417
28	130	6.5	792	2	A84308
29	129.5	6.5	403	2	B83408

30	128.5	6.5	183	2	A72499	hypothetical prote
31	127.5	6.4	288	2	S18438	sporulation protei
32	127.5	6.4	618	2	B87566	voltage gated chlo
33	125.5	6.3	125	2	B87343	ABC transporter, A
34	125.5	6.3	840	2	F84213	hypothetical prote
35	125	6.3	404	2	G70886	probable integral
36	124.5	6.3	378	2	H75213	serine proteinase
37	124.5	6.3	755	2	B75346	probable competent
38	124	6.2	285	2	T34615	NADH2 dehydrogenas
39	123.5	6.2	367	2	B70839	hypothetical prote
40	123	6.2	370	2	H84351	hypothetical prote
41	122.5	6.2	219	1	D70057	conserved hypotnet
42	122	6.1	1029	1	D83120	probable RND efflu
43	121.5	6.1	347	2	A95849	probable ABC trans
44	121	6.1	217	1	G70459	conserved hypotnet
45	121	6.1	709	2	F75584	hypothetical prote

## ALIGNMENTS

RESULT 1									
H70572	hypothetical protein RV2625c - Mycobacterium tuberculosis (strain H37RV)								
C:Species: Mycobacterium tuberculosis									
C:Date: 17-Jul-1998	#sequence_revision 17-Jul-1998	#text_change 09-Jul-2004							
C:Accession: H70572									
R:Coile, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Genicels, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.									
Nature 393, 537-544, 1998									
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.									
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome									
A:Reference number: A70500; MUID:98295987; PMID:9634230									
A:Accession: H70572									
A>Status: preliminary; nucleic acid sequence not shown; translation not shown									
A:Molecule type: DNA									
A:Residues: 1-393 <COL>									
A:Cross-references: UNIPROT:O06187; UNIPARC:UPI00000D1035; GB:Z95387; GB:ALJ23456; NID:9									
A:Experimental source: strain H37RV									
C:Gene: RV2625c									
C:Superfamily: conserved hypothetical protein M0392; CBS homology									
Query Match	100.0%	Score 1990:	DB 2:	Length 393:					
Query Similarity	100.0%	Pred. No. 2.6e-137:							
Matches 393:	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;					
QY	1	MRD1PILGR1AGFVVNVHMSVLTLMFTWSLATMLPCTVGYPVAVVYMLGAGAVMLL	60						
DB	1	MRD1PILGR1AGFVVNVHMSVLTLMFTWSLATMLPCTVGYPVAVVYMLGAGAVMLL	60						
QY	61	ASLAEHIAHVVARRAGVSVEVTLMFGVTLAGEAKPKKAFRAFPAGPATSLALS	120						
DB	61	ASLAEHIAHVVARRAGVSVEVTLMFGVTLAGEAKPKKAFRAFPAGPATSLALS	120						
QY	121	ATFGALATTLTAGVTPPAIVSVAMWLTAVNLLGIFNLPGAPLDGGTLVAAYMRHGD	180						
DB	121	ATFGALATTLTAGVTPPAIVSVAMWLTAVNLLGIFNLPGAPLDGGTLVAAYMRHGD	180						
QY	181	SVRAGIGARAGRVAVLALGLAEFVAGGLVGSVWMLAFTGWFIPAAAREEETRISTQQ	240						
DB	181	SVRAGIGARAGRVAVLALGLAEFVAGGLVGSVWMLAFTGWFIPAAAREEETRISTQQ	240						
QY	241	LFAGRVADANTAOPTAPGAINVEDFQRYVLGERSHAYVADRDGSTITGLVALRQIRD	300						
DB	241	LFAGRVADANTAOPTAPGAINVEDFQRYVLGERSHAYVADRDGSTITGLVALRQIRD	300						
QY	301	VAPSRSTSVGDIAPLPHSVPTARPOEPLTALLERMAPLGPBRRAVTEGSAVVGITP	360						
DB	301	VAPSRSTSVGDIAPLPHSVPTARPOEPLTALLERMAPLGPBRRAVTEGSAVVGITP	360						
QY	361	SDVARLIDVRLAQPETFTTSPODADRFS DAG 393							



Db 361 SDVARLIDVRLAQBEPFTTSPQADRFSDAG 393  
|||||  
RESULT 2  
AF2036  
hypothetical protein al1844 [imported] - Noctoc sp. (strain PCC 7120)  
C/Species: Noctoc sp. PCC 7120  
A/Note: Noctoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C/Accession: AF2036  
R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasa moto, S.; Watanabe, A.; Itiguchi, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A/Reference number: AB1807; MUID:21595285; PMID:11759840  
A/Accession: AF2036  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-389 <KIR>  
A/Cross-references: UNIPROT:O8VYX5; UNIPARC:UPI00000CE21E; GB:BA000019; PIDN:BA073543.1;  
A/Experimental source: strain PCC 7120  
C/Genetics:  
A/Gene: al1844  
C/Superfamily: conserved hypothetical protein M0392; CBS homology  
Query Match 24.4%; Score 486.5; DB 2; Length 389;  
Best Local Similarity 30.7%; Pred. No. 5e-28;  
Matches 118; Conservative 78; Mismatches 169; Indels 19; Gaps 7;  
Qy 1 MRDAIPGRIAGFVNVNWSVLVILMFTWSLATMLPGTVGGYP-----AVVYWLGGAG 54  
Db 17 MNGTIRAGNLFQIPFYIHPSWFLVGLVTWYS--GGLIABFQSLGRLAVV--LGIA 70  
Qy 55 GAVMLASLHSLAHLAAVVARAGVSVSVTLMFGVTALGGEKTKAARLAFAPRA 114  
Db 71 TALLFASVVAHEHSHSFVAIRQGINVSTLTFGGIASHSEKSKTPAGAFVWAIAGPL 130  
Qy 115 TSLASATFGALATTLGAVTRPAIVISVAMWLATVNLGLENLPGALGGRLVRAVL 174  
Db 131 VSLICGTVTTIGTAVTAVTGLAAILGV---LASVNLALFNLIPGLPLOGNITLRAIV 187  
Qy 175 WRRHGDVSRVAGIGARAGRVVALVILALGLAEFVAGLVGGVWLAFICWTFPAAREEET 234  
Db 188 WKVTGNPFKGVTPASRVGQIRGVWVAIGIPILYFGSFANVMNLLIGFLLQVAGNAQA 247  
Qy 235 RISTQQLFAGRVADAMTAQHTAPGWINVEDFIQRYVL-GERHSAIVVAARDGSITGLV 293  
Db 248 FARVOEKLTGLTADAVTTDSPIVSAHLSREFFADDOITIQGQNRFLVTNNAQGLVGA 307  
Qy 294 ALRQLRDVAPSRSTTSVGDIALPLHSPVTPARPOEPLTALLERMAPLGPSPRSALVTESGA 353  
Db 308 ALDDLRNIPITSWETQIQQWRRPIQST-TIKSSQPLLEVVQLEQQLSLPYILDNGV 366  
Qy 354 VVGIVTPESDVARLIDVYRLAQPEP 377  
Db 367 LIGILKRAIITQLQ--NGTQPNP 388  
RESULT 3  
AE2293  
hypothetical protein alr3900 [imported] - Noctoc sp. (strain PCC 7120)  
C/Species: Noctoc sp. PCC 7120  
A/Note: Noctoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C/Accession: AE2293  
R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasa moto, S.; Watanabe, A.; Itiguchi, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A/Reference number: AB1807; MUID:21595285; PMID:11759840  
A/Accession: AE2293  
A/Status: preliminary

A/Molecule type: DNA  
A/Residues: 1-399 <KIR>  
A/Cross-references: UNIPROT:Q8YQD3; UNIPARC:UPI00000CE945; GB:BA000019; PIDN:BA075599.1;  
A/Experimental source: strain PCC 7120  
C/Genetics:  
A/Gene: alr3900  
C/Superfamily: conserved hypothetical protein M0392; CBS homology  
Query Match 24.0%; Score 477; DB 2; Length 399;  
Best Local Similarity 34.1%; Pred. No. 2.5e-27;  
Matches 120; Conservative 59; Mismatches 163; Indels 10; Gaps 4;  
Qy 25 LWFETWSLATMLPGTV-GGYPVVYVWLGGAGVAMLIASLHSLAAVVARAGVSVES 83  
Db 21 LMFYILGLATLINFVAOEWGTVTAMTAGIMSLILGSGVSLHSLHSLARSQGIYVNS 80  
Qy 84 VTLMLPGVVALGGEATTPKAPRIAPAGPATSLSATGALATITAGVTPAIVSVA 143  
Db 81 ITLFLPGIATIEESTTPGAKFQVALAGPLVSTGL--FULLRSGTVVSDSPVSMNV 137  
Qy 144 WMLATVNLGLFNLPLGAPLDGRLVRAVLMRRHGDVSRVAGIGARAGRVVALVILALG 203  
Db 138 GDLARINLVVALFNLIPGLPLOGQVLLKALMQITGDRPQAVHMAAKAGQILGICALAG 197  
Qy 204 LA-EFVAGLVGGVWLAFICWTFPAAREESTRISTQQLFAGRVADAMTAQHTAPGWI 262  
Db 198 FAVDFPRELVATGIMALLGFGVRNANSYDRVTTLQETLLEVRKAVDAMTRDFEVIDANQ 257  
Qy 263 NYEDFIQRYVLGERHSAIVPADRDGSITGLVALRQLDVAIPSRSTTSVGDIALPLHSPV 322  
Db 258 TLRSFADSYLANSPEVYTPASDGRYRGWALDELVERSEWETQLSHIAHPLTEIP 317  
Qy 323 TARQEBPLTALLERMAPLGPSPRSALVTESGAVVGVITPSPDVARLIDVYRLAQ 374  
Db 318 TVAESTIAGVINKLENQLEPRVTVLTPAGAVAGIIDRGDI-----VGLAQ 364  
RESULT 4  
D84367  
hypothetical protein Vng2168c [imported] - Halobacterium sp. NRC-1  
C/Species: Halobacterium sp. NRC-1  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C/Accession: D84367  
R/Ng, W.V.; Kennedy, S.P.; Mahiras, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S  
; Leithauser, B.; Keller, K.; Cruz, R.; Danon, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor  
Jung, K.H.; Alam, M.; Freilich, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A/Authors: Hou, S.; Daniel, C.J.; Demito, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li  
; Leithauser, B.; Keller, K.; Cruz, R.; Danon, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor  
Jung, K.H.; Alam, M.; Freilich, T.  
A/Title: Genome sequence of Halobacterium species NRC-1.  
A/Reference number: A84160; MUID:20504483; PMID:11016950  
A/Accession: D84367  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-390 <STO>  
A/Cross-references: UNIPROT:Q9HNB6; UNIPARC:UPI0000063A7B; GB:AB004437; NID:g10581585; P  
C/Genetics:  
A/Gene: VNG2168C  
C/Superfamily: conserved hypothetical protein M0392; CBS homology  
Query Match 23.1%; Score 459; DB 2; Length 390;  
Best Local Similarity 33.9%; Pred. No. 5e-26;  
Matches 132; Conservative 63; Mismatches 152; Indels 42; Gaps 14;  
Qy 7 LGRAGFVNVNWSVLVILMFTWSLATMLP-----GTGQGYP-----AVVYWL 51  
Db 6 IGRVFGPIPKIDVTFLVLPVFAVILGTGLEWVSTLANGAPFAAGLADATALTAGSVKVL 65  
Qy 52 GAGGAVMLASLHSLAHLAAVVARAGVSVSVTLMFGVTALGGEKTKAARLAFAPRA 111  
Db 66 GAATAAGVLFQVVLHSHSVAMRWYDFITDAITLWLVGGVASTLPDPEEDQELCTIALA 125  
Qy 112 GPATSLASLA-TFGALITTLGAVTRPAIVISVAMWLATVNLGLENLPGALGGRLV 170

Db 126 GPANVALAAVAYGLAALPSSLDVTRVRFG---YLALMNLALAAFNLLPGFPMDCGRVL 182  
Qy 171 RAYLWRRGDSVYRAGIGARAGRVVALLALGLAEFVAGLVGGVWMLFGMFIFAAR 230  
Db 183 RALLARTN-TPARAVAKIAAEVKGKLPFVLGAGLLSF-----NVILIGVAFYIGAA 234  
Qy 231 EEFRTISIQQLFAGRVADAMT--AQPTAPGMINVEPFIQRYVLGERHSAYPVADRGS 288  
Db 235 GEAQRTANMAAFEGVTADVMTVPASDVHTVATASVADLMDS-MLEHRTGYPVF-RDAT 292  
Qy 289 ITGLALAEQLRDVAPRSRSTTSVGD1-ALPLHSVP-TARPOEPLALLERMAPLGRPSRA 346  
Db 293 AVGMTTLDDARSVRAVERDAMRVADWSDDYITLPGSDATDALDQLQEHSV-----GRL 347  
Qy 347 LVTEGSA-VGQIVTPSPDVARLIDVYRLAQ 374  
Db 348 LVVDADGEMVGLITRSD---LMDAFGIQ 373

## RESULT 5

hypochemical protein - *Synechocystis* sp. (strain PCC 6803)  
C/Species: *Synechocystis* sp.  
C/Variety: PCC 6803  
C/Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C/Accession: S76029  
R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Matenabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-116, 1996  
A/Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
8.  
A/Reference number: S74322; MUID:97061201; PMID:8905231  
A/Accession: S76029  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-379 <KAN>  
A/Cross-references: UNIPROT:O55518; UNIPARC:UPI0000139445; EMBL:D64006; GB:AB001339; NID  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C/Superfamily: conserved hypothetical protein M0392; CBS homology

Query Match 21.4%; Score 425.5; DB 2; Length 379;  
Beet Local Similarity 30.0%; Pred. No. 1.3e-23;  
Matches 115; Conservative 70; Mismatches 147; Indels 51; Gaps 11;  
Qy 1 MRDAIPGRING-FVNVHMSVLIWLFMTSLATML---PGTVGYPVAVVWMLGAG 55  
Db 11 MNNNIRVSGSLGIPYVPSW--FLLGLVLTSLSGDLARFPQSGGTP---WILGLIT 64  
Qy 56 AVMLIASILAEHLAAVAVRAGVSVESVTMLFGVTVLAGEAKTPKAPRIAPAGPAT 115  
Db 65 ALLLPASVVAHELGHSVALAQIEVKSITLFLFGGLASLEKESWTPMQAFVAVALAGPAV 124  
Qy 116 SLAISATFGALAITLAGRTPAIV--ISVAMWLATVNLGLFNLLPGAPLDGGLVAY 173  
Db 125 SLVL-----FLGLITVGTQIPLPVGQAIIIGLGMINIALLFNLIPLDGGVNLKSI 179  
Qy 174 LMRHGDGSRVAGIGARAGRVVALLALGLAEFVAGLVGGVWMLFGMFIFAAREE 233  
Db 180 VMQITGNONKGLILASRVGGEGMIALIGSLGIMNIPISFWTLLIGMFLONAGSSA 239  
Qy 234 TRISTQQLFAGRVADAMTAPHTAPGMINVEPFIQRYVLGER-HSAYPVADRGSITGL 292  
Db 240 RNAQVKEGMEAFTEADAVIPNSPIPAGLNREFANDVIGKTPMRRLVLGADNQLLGV 299  
Qy 293 VALKQLRDVAPRSRSTTSVGDIALPLHSVPTAR-PQELTALLERMAPLGRPSALVTEG 351  
Db 300 LATBIRK-----HVPTSDWPQVTVDSIMQY-----PQQWTVVAN 334  
Qy 352 SAVVGIVTPSPDVARLIDVYRLAQ 374  
Db 335 QSLF-----EVAQLLDDQKLS 351

## RESULT 6

D69291  
conserved hypothetical protein AF0332 - *Archaeoglobus fulgidus*  
C/Species: *Archaeoglobus fulgidus*  
C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C/Accession: D69291  
R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kyrnes, E.F.  
Glock, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A/Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes, S.  
Smith, H.O.; Weese, C.R.; Venter, J.C.  
A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
A/Reference number: A69250; MUID:98049343; PMID:9389475  
A/Accession: D69291  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-362 <KLE>  
A/Cross-references: UNIPROT:O29915; UNIPARC:UPI000005715F; GB:AE001082; GB:AE000782; NID  
C/Superfamily: conserved hypothetical protein M0392; CBS homology  
F.299-346/Domain: CBS homology <CBS>

Query Match 21.0%; Score 418; DB 2; Length 362;  
Beet Local Similarity 30.5%; Pred. No. 4.4e-23;  
Matches 113; Conservative 70; Mismatches 162; Indels 26; Gaps 10;  
Qy 1 MRDAIPGRINGFVNVHMSVLIWLFMTSLA-TMLPGTVGYPVAVVWMLGAGAVML 59  
Db 1 MRASFKIFRVGIDIVEHISLILVLIYAFVSPPPYGAPNPLSERIILSSMAAVGL 60  
Qy 60 LASILAEHLAAVAVRAGVSVESVTMLFGVTVLAGEAKTPKAPRIAPAGPATSLAL 119  
Db 61 FASILAEHLGHSIVARRVGRIRGIMLFIQGVAMQELPKKPEERLVVAISGATS--- 117  
Qy 120 SATFGALAITLAGRTPAIVSVAMWLTATVNLGLFNLLPGAPLDGGRVAVYLRBH 178  
Db 118 ---GIAVNSALSSIPVABISAFLLPGYUNFLIALFNLLPAPMDGGRVLSFLAKR 174  
Qy 179 GDSVRAGIGARAGRVVALLALGLAEFVAGLVGGVWMLFGMFIFAAREESTIST 238  
Db 175 SYAATNI-MAEIORALAIMAIF-----GIFTPMILILALFYIGANEERVLVL 225  
Qy 239 QQLFAGRVADAMTAPHTAPGMINVEPFIQRYVLGERHSAYPVADRGSITGLVALRQL 298  
Db 226 ENVIGRVADVMTTEVVTVPENTVSEVID-LILKTKHIGFPVBESE-RLVGIITLHDI 283  
Qy 299 RDVAPRSRSTTSVGDIALPLHSVPTARPOEPLALLERMAPLGRPSALVTEGSAVYIV 358  
Db 284 IGVSPEER---VGNIN--GREVVAVSPNQSAPFAFKIMSEMG-IGRLPVVEHGRVVGIV 336  
Qy 359 TPSPDVARLIDV 369  
Db 337 SRSDLMRIKEI 347

## RESULT 7

A69209  
conserved hypothetical protein MTH816 - *Methanobacterium thermoautotrophicum* (strain Deli  
C/Species: *Methanobacterium thermoautotrophicum*  
C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C/Accession: A69209  
R/Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
J. Qiu, D.; Spadafora, R.; Viscare, R.; Wang, Y.; Mierboweki, J.; Gibson, R.; Jivani, N.  
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Neelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A/Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: func.  
A/Reference number: A69000; MUID:98037514; PMID:93711463  
A/Accession: A69209  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-341 <MTH>  
A/Cross-references: UNIPROT:O26907; UNIPARC:UPI0000066737; GB:AE000859; GB:AE000666; NID  
A/Experimental source: strain Delta H

C;Geneticbg:  
A;Gene: MTH916  
C;Superfamily: conserved hypothetical protein M0392; CBS homology Ff287-334/Domain: CBS homology <CBS>  
  
Query Match 19.0%; Score 377.5; DB 2; Length 341;  
Beat Local Similarity 30.2%; Pred. No. 3.6e-20;  
Matches 111; Conservative 74; Mismatches 147; Indels 35; Gaps 13;  
  
Qy 4 AIPGRIAGFVVNVMSVLVILMLFTWSLATMTPGVGGYPAVVYMLLAGCAVMLLASL 63  
Db 3 SVKIFSVSGTIEIDFSFL-LLMFLITYLATL-----GFISVNLAVL----ITLVPTV 51  
Qy 64 LAHELAHVAVARRAVSVEVTLMFGSVALGGEAKTPKAFFIAFGPATSLALSATF 123  
Db 52 VIHLEAHSYVALRFCKVKISILLPIGSVMREIPIRIPQEFMIISIAGPLTI-LMALI 110  
Qy 124 GALATTLAGVTRPALIVSVAMMATVNILLGLFNILPCAPADCGRLRVAYIMRHGDSVR 183  
Db 111 TAVPL-LFGLKGAA--ASFADDFLAVNLLLAIFNIIPAFPDGKRIRAIATAEL-VYIR 166  
Qy 184 AGIGAARAGRVAVALTLALGLAEFVAGLGCVGMALFAFMFIIPAAREETRISTOOLFA 243  
Db 167 ATRIASNIGKIITAVM-----AVTGLPFNFFILLGIPTIYGAEDEYATLISILLE 218  
Qy 244 GVRVADMTAQPHRTAPGWINVEDFIQRYYLGERHSAPVADRDSTIGLVALROLRDVAP 303  
Db 219 GVRVADVNTERTPVTLHPHTVKEALD-VMFPEKHMGVETEA-GELRGIVTFHDISDASR 276  
Qy 304 SRR-STTSVGIALPLHSVPFPARPOEPLITALLERKAPGPSRALYTEGSAVNGITVPSD 362  
Db 277 DLKREDVWTVGTGVV-----TVRDBEEVTGALKKKNNRL-QLGRLPVMDGKLGTIIISRTD 328  
Qy 363 VARLDV 369  
Db 329 IVRTLNL 335

RESULT 8  
H64348  
hypothetical protein homolog M0392 - Methanococcus jannaschii  
C;Species: Methanococcus jannaschii  
C;Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jul-2004  
C;Accession: H64348  
R;Butl, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
R.; Reich, C.I.; Overbeek, R.; Kikunas, E.F.; Weissstock, K.G.; Merrick, J.M.; Glodek, A.;  
Tron, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurec, M.A.  
Science 273, 1058-1073, 1996  
A;Authors: Kalne, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.  
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
A;Reference number: A64300; MUID:96337999; PMID:8688087  
A;Accession: H64348  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Cross-references: UNIPROT:Q57837; UNIPARC:UPI000013972E; GB:U67492; GB:L77117; NID:g15  
C;Genetics:  
C;Map position: FOR355789-356808  
C;Superfamily: conserved hypothetical protein M0392; CBS homology  
  
Query Match 16.5%; Score 328.5; DB 2; Length 339;  
Beat Local Similarity 24.7%; Pred. No. 1.3e-16;  
Matches 95; Conservative 86; Mismatches 144; Indels 59; Gaps 14;  
  
Qy 1 MRDAIPGRIAGFVVNVHMSVLVILMLFTWSLATMTPGVGGYPAVVYMLLAGCAVMLL 60  
Db 1 MNYSIRLFKINGIPDELH--ITPILFLVIIIGLSIMNSI--FNAVLPF-----ILLF 48  
Qy 61 ASLAHELAHVAVARRAVSVEVTLMFGSVALGGEAKTPKAFFIAFGPATSLALSATF 119  
Db 49 VSVVAHELGHSHYVAKRGVKTLEKITLLPIGGVAMND--KIPGEGELIGTAGLVSFTI 105  
Qy 120 SATGCALA-----ITLAGVTRPALIVSVAMMATVNILLGLFNILPCAPADCGRLRVAYIM 175

Db 106 GIVLLIVSQFDINING-----YPLVYTLSTLNMVLGGFNLIIPAFMDGGKILRAILS 158  
QY 176 RHHDGSVAGIGCAABAGRVVALVTLALGLAEFVAGLVGVWVLAFIGWPIFAAAREBETR 235  
Db 159 KKYGLKSKSTKI--AANIIGSLALIMLFLGL-----LSNMIIILVLSFYFGAEQESRV 210  
QY 236 ISTQOLFAGVRVADAMTAQPHTPAGMINVEDFIQRYVLGERSAIVYADBDGSIITGLVAL 295  
Db 211 VEVETIFPNKIKKDKIMTPNPVVTDPMSIEEFID--FMLKHKVFYGVVE--NGRLVGCIGI 268  
QY 286 RQLRVAASRSRSTSVGDALPLHVSPTARPOEPL-----TLLEEMAPLGRPSRALYTE 350  
Db 269 GNI-----HKKEGVTRDYM-----EKPVVSEDTDIKELRKXANTDRVFVE 311  
QY 351 GSAVGIPTSPDVARLIDVYRLAQ 374  
Db 312 GGLKGLITSKDIDILPAMSTLELKE 335

RESULT 9  
H95972  
conserved hypotheical membrane protein SMD20925 [imported] - Sinorhizobium meliloti (ser:  
C:Species: Sinorhizobium meliloti  
C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: H95972  
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernant  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo  
A:Reference number: A95842; MUID:21396508; PMID:11481431  
A:Accession: H95972  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-372 <KDS>  
A:Cross-references: UNIPROT:Q92UP7; UNIPARC:UPI00000CB741; GB:AL51985; PIDN:CA649448.1;  
A:Residues: 1-372 <KDS>  
A:Experimental source: strain 1021, megaplasmid pSymB  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Punler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federle, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiser, E.; Komp, C.; Lelaure,  
hebaull, P.; Vandembol, N.; Vorholter, F.J.; Weidner, S.; Waller, D.H.; Wong, K.; Yeh, K.C.  
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: SMD20925  
A:Genome: plasmid  
C:Superfamily: conserved hypotheical protein M0392; CBS homology

Query Match 16.5%; Score 328; DB 2; Length 372;  
Best Local Similarity 29.4%; Pred. No. 1.6e-16;  
Matches 113; Conservative 66; Mismatches 156; Indels 50; Gaps 17;

Db 7 LGRIAGFVNNVMS--VLVILMFTWSLATMLDPTGVGYPAVYVYLLGAGAVMLASLLA 65  
QY 7 IGTAGTARVHVTFALLIWI--W---LMHYRIGGTBPA--MEGIATVAVFVCVL 57  
Db 66 HELAAVAVARRGVSVESTTLMFGVTALGGEAKTPKAARFAPAGPATSLASRFGA 125  
QY 58 HEFGIHAARRGIGITPTDITLLPIGGVARLERMPERPGEFVAIAGPLVNAIAAVALIA 117  
Db 126 LAIT-----LAGVTPATVYSVMMVLAATNLLGLFNLLPGAPLDGGRVAVYLMRRHD 180  
QY 118 ILGGSGEQIACVEDPQSFAR--LAGVNVFLVITNMIPAFMDGGRVVALAASR-- 172  
Db 181 SVRACIGARAGRVVALVTLALGLAEFVAG--GLVGVWVLAFIGWPIFAAAREBETR 238  
QY 173 -----LMSRATQIAA--TIGGLA-FVFGFGLFNPPLIFAIIVYLAATDAQNAQI 224  
Db 239 QQLFAGVRVADAMTAQPHTPAGMINVEDFIQRYVLGERSAIVYADBDGSIITGLVA--- 294  
QY 225 REIGSGWISDWMITEFATLDRSATIDDAIDT--LATTORPEFVVDAAHGFEGLLTRDDM 283



	Best Local Similarity	25.9%; Pred No. 0.0011;	Mismatches	80;	Indels	37;	Gaps	7
	Matches	53;	Conservative	35;				
OY	6 VVTWLGAGGAVMLASL-LAHELAHAVAPRACGSVESVTTMTMFGGTALGG-EAKTPK	103	:	:	:	:	:	:
Dd	13 ILLEFIIGRGELTFESVLIVIHFEFHYLIARNFSGSFDIKILPGCAVIRLKRIYAQOK	72	:	:	:	:	:	:
OY	104 AAFRAIFGPATSLIASITFCALATIAGVRPAIVI-----SVAMMATVNLLGLFN	157	:	:	:	:	:	:
Dd	73 EDIITSLSGPLNTLTIAVF-----VYIFNDSGRHIYSLSYPNTLSGLFN	118	:	:	:	:	:	:
OY	158 LLPGAFLDGGRLVRYLYIMRRHGDVSBRAGIGAARSRVALYLIALGLEFAVGAGLVG----	214	:	:	:	:	:	:
Dd	119 LIPAFPLOGCRMRMKVELSEK-----LPKNABEVSTIYSTIVIGII-FCIFSIVGLELF	169	:	:	:	:	:	:
OY	215 ---GVMLAFICMFIPAAREETTRI	236	:	:	:	:	:	:
Dd	170 SVKSINMLVVAAFFILFCSPKEKGKM	194	:	:	:	:	:	:

## RESULT 13

hypothetical protein *alr2114* [imported] - Noctoc sp. (strain PCC 7120)  
A:Note: Noctoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120  
C:Species: Noctoc sp. strain PCC 7120  
C:Accession: AD20701 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C:Date: 14-Dec-2001  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kurita, T.; Sasaoka, S.; Matsumoto, A.; Iriyachanok, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.; Nakazaki, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete genomic sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena*  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AD2070  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-493 <KUR>  
A:Cross-references: UNIPROT:O8YV67; UNIPARC:UPI000000CE30E; GB:BA000019, PIDN:BAJ73813.1  
A:Experimental source: strain PCC 7120  
C:Genetics:  
C:Gene: *alr2114*

	Query Match	Similarity	7.2%	Score 144;	DB 2;	Length 493;
	Best Local	Similarity	25.4%;	Pred. No. 0.0051;	Mismatches 65;	Gaps 8
	Matches	Conservative	33;	Pid.	Indels 52;	Gaps 8
Oy	44	PAVVVWLLGAGGAVMLLSILNHEIAHNVVARAGSVY-----SVTLMLEGVTALGG	97			
Dd	245	PARIGDALPIGIGIVL--LVNHEIGHMLVARRHQVRISWPFPLDAVOIGSGAITRFES	302			
Oy	98	EAKTPKAAFLRLAFAPD---ATSLA-----SATFGALATLT	130			
Dd	303	LLEPNRSVLFDLIAVAPRIAGTVSLLMTLVGLLLSHQGSIFQLPNOFPQGISIIIVGSLARV	362			
Oy	131	AG--VRTPAI-----VISVAAMLATYNLLLGLENLPFGARDGGRVRAVLRHRHDSVRA	184			
Dd	363	LGSAIAGSLVNVNHLVIITGM-----LGVITANLMPRAGSLDGRIITAIYGRKTA-----	413			
Oy	185	GIGAARAGRVALVTLIALGLA	205			
Dd	414	--GRATPATLIVLAVLSIGNA	432			

## RESULT 14

U1142  
U1143  
U1144  
U1145  
U1146  
U1147  
U1148  
U1149  
U1150  
U1151  
U1152  
U1153  
U1154  
U1155  
U1156  
U1157  
U1158  
U1159  
U1160  
U1161  
U1162  
U1163  
U1164  
U1165  
U1166  
U1167  
U1168  
U1169  
U1170  
U1171  
U1172  
U1173  
U1174  
U1175  
U1176  
U1177  
U1178  
U1179  
U1180  
U1181  
U1182  
U1183  
U1184  
U1185  
U1186  
U1187  
U1188  
U1189  
U1190  
U1191  
U1192  
U1193  
U1194  
U1195  
U1196  
U1197  
U1198  
U1199  
U1200  
U1201  
U1202  
U1203  
U1204  
U1205  
U1206  
U1207  
U1208  
U1209  
U1210  
U1211  
U1212  
U1213  
U1214  
U1215  
U1216  
U1217  
U1218  
U1219  
U1220  
U1221  
U1222  
U1223  
U1224  
U1225  
U1226  
U1227  
U1228  
U1229  
U1230  
U1231  
U1232  
U1233  
U1234  
U1235  
U1236  
U1237  
U1238  
U1239  
U1240  
U1241  
U1242  
U1243  
U1244  
U1245  
U1246  
U1247  
U1248  
U1249  
U1250  
U1251  
U1252  
U1253  
U1254  
U1255  
U1256  
U1257  
U1258  
U1259  
U1260  
U1261  
U1262  
U1263  
U1264  
U1265  
U1266  
U1267  
U1268  
U1269  
U1270  
U1271  
U1272  
U1273  
U1274  
U1275  
U1276  
U1277  
U1278  
U1279  
U1280  
U1281  
U1282  
U1283  
U1284  
U1285  
U1286  
U1287  
U1288  
U1289  
U1290  
U1291  
U1292  
U1293  
U1294  
U1295  
U1296  
U1297  
U1298  
U1299  
U1300  
U1301  
U1302  
U1303  
U1304  
U1305  
U1306  
U1307  
U1308  
U1309  
U1310  
U1311  
U1312  
U1313  
U1314  
U1315  
U1316  
U1317  
U1318  
U1319  
U1320  
U1321  
U1322  
U1323  
U1324  
U1325  
U1326  
U1327  
U1328  
U1329  
U1330  
U1331  
U1332  
U1333  
U1334  
U1335  
U1336  
U1337  
U1338  
U1339  
U1340  
U1341  
U1342  
U1343  
U1344  
U1345  
U1346  
U1347  
U1348  
U1349  
U1350  
U1351  
U1352  
U1353  
U1354  
U1355  
U1356  
U1357  
U1358  
U1359  
U1360  
U1361  
U1362  
U1363  
U1364  
U1365  
U1366  
U1367  
U1368  
U1369  
U1370  
U1371  
U1372  
U1373  
U1374  
U1375  
U1376  
U1377  
U1378  
U1379  
U1380  
U1381  
U1382  
U1383  
U1384  
U1385  
U1386  
U1387  
U1388  
U1389  
U1390  
U1391  
U1392  
U1393  
U1394  
U1395  
U1396  
U1397  
U1398  
U1399  
U1400  
U1401  
U1402  
U1403  
U1404  
U1405  
U1406  
U1407  
U1408  
U1409  
U1410  
U1411  
U1412  
U1413  
U1414  
U1415  
U1416  
U1417  
U1418  
U1419  
U1420  
U1421  
U1422  
U1423  
U1424  
U1425  
U1426  
U1427  
U1428  
U1429  
U1430  
U1431  
U1432  
U1433  
U1434  
U1435  
U1436  
U1437  
U1438  
U1439  
U1440  
U1441  
U1442  
U1443  
U1444  
U1445  
U1446  
U1447  
U1448  
U1449  
U1450  
U1451  
U1452  
U1453  
U1454  
U1455  
U1456  
U1457  
U1458  
U1459  
U1460  
U1461  
U1462  
U1463  
U1464  
U1465  
U1466  
U1467  
U1468  
U1469  
U1470  
U1471  
U1472  
U1473  
U1474  
U1475  
U1476  
U1477  
U1478  
U1479  
U1480  
U1481  
U1482  
U1483  
U1484  
U1485  
U1486  
U1487  
U1488  
U1489  
U1490  
U1491  
U1492  
U1493  
U1494  
U1495  
U1496  
U1497  
U1498  
U1499  
U1500  
U1501  
U1502  
U1503  
U1504  
U1505  
U1506  
U1507  
U1508  
U1509  
U1510  
U1511  
U1512  
U1513  
U1514  
U1515  
U1516  
U1517  
U1518  
U1519  
U1520  
U1521  
U1522  
U1523  
U1524  
U1525  
U1526  
U1527  
U1528  
U1529  
U1530  
U1531  
U1532  
U1533  
U1534  
U1535  
U1536  
U1537  
U1538  
U1539  
U1540  
U1541  
U1542  
U1543  
U1544  
U1545  
U1546  
U1547  
U1548  
U1549  
U1550  
U1551  
U1552  
U1553  
U1554  
U1555  
U1556  
U1557  
U1558  
U1559  
U1560  
U1561  
U1562  
U1563  
U1564  
U1565  
U1566  
U1567  
U1568  
U1569  
U1570  
U1571  
U1572  
U1573  
U1574  
U1575  
U1576  
U1577  
U1578  
U1579  
U1580  
U1581  
U1582  
U1583  
U1584  
U1585  
U1586  
U1587  
U1588  
U1589  
U1590  
U1591  
U1592  
U1593  
U1594  
U1595  
U1596  
U1597  
U1598  
U1599  
U1600  
U1601  
U1602  
U1603  
U1604  
U1605  
U1606  
U1607  
U1608  
U1609  
U1610  
U1611  
U1612  
U1613  
U1614  
U1615  
U1616  
U1617  
U1618  
U1619  
U1620  
U1621  
U1622  
U1623  
U1624  
U1625  
U1626  
U1627  
U1628  
U1629  
U1630  
U1631  
U1632  
U1633  
U1634  
U1635  
U1636  
U1637  
U1638  
U1639  
U1640  
U1641  
U1642  
U1643  
U1644  
U1645  
U1646  
U1647  
U1648  
U1649  
U1650  
U1651  
U1652  
U1653  
U1654  
U1655  
U1656  
U1657  
U1658  
U1659  
U1660  
U1661  
U1662  
U1663  
U1664  
U1665  
U1666  
U1667  
U1668  
U1669  
U1670  
U1671  
U1672  
U1673  
U1674  
U1675  
U1676  
U1677  
U1678  
U1679  
U1680  
U1681  
U1682  
U1683  
U1684  
U1685  
U1686  
U1687  
U1688  
U1689  
U1690  
U1691  
U1692  
U1693  
U1694  
U1695  
U1696  
U1697  
U1698  
U1699  
U1700  
U1701  
U1702  
U1703  
U1704  
U1705  
U1706  
U1707  
U1708  
U1709  
U1710  
U1711  
U1712  
U1713  
U1714  
U1715  
U1716  
U1717  
U1718  
U1719  
U1720  
U1721  
U1722  
U1723  
U1724  
U1725  
U1726

A:Accession: D71142  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-573 <KAM>  
A:Cross-references: UNIPROT:O50809; UNIPARC:UPI0000062DB7; GB:AF000002; NID:g3235129; PBI:A8Experimental source: strain O73  
A>Note: This accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH0351

Query Match	7.2%	Score 143.5	DB 2	Length 373
Best Local Similarity	22.0%	Pred. No. 0.004		
Matches	66	Conservative	37	Mismatches 92
				Indels 105
				Gaps 9
QY	6	PLGRINGPVNNVNRSLVLITMLEFTWSLAKTM-----LPGTVGGYPAVYVWLLGAGGAVWLLA	61	
DB	89	PLGLITLFPVLTLESTFEPAGYILSSLYVATLWELNLPQIKVNYLNALNFSIG-----TIS	142	
QY	62	SLLAHELAAHVAVRRAGVSVE-----SVTLMLFGGVTALGGEAKTPKAAFRFAEGPA	114	
DB	143	ILGTHGKHAKIANTLHNVNSTFPYITPPSPFGLTGAVIRKSEPIPRNAAVDLGAAGPI	202	
QY	115	TSIALSATFGAALITLAVGRTPAIV-----	139	
DB	203	AGLLV-----ALPVTIIIGLKSIVIPVDYLKQGETIYRGTSILEFYALTKFVLGNLPQSG	257	
QY	140	-----ISVAMWLTATVNLILGLFNLLPGAPLDGGRLVRAVYLMRRHGDSSVRAGIGAAAGRV	194	
DB	258	ILHPPLAVAGW---VGILVTFPLNLLIPAAOLDGSHIRALM-----PERAHRI	301	
QY	195	V--ALVILNLAGLAEFPAGGLVGGWLLAFIG-----WPIFA	227	
DB	302	LTVAGGITLTGASYPMPGWLWGLIIILLMGRIGNPGALDEVTPLTTPGRKALAILIWTIFA	361	

## RESULT 15

B/5010  
 Hypothetical protein PAB1063 - *Pyrococcus abyssi* (strain Oresay)  
 C:Species: *Pyrococcus abyssi*  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004  
 C:Accession: B75010  
 R:Anonymous, Genoscope  
 submitted to the EMBL Data Library, July 1999  
 A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome structure  
 A:Reference number: A75001  
 A:Accession: B75010  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-409 <KAW>  
 A:Cross-references: UNIPROT:O9UY92; UNIPARC:UP10000034560; GB:AJ248288; GB:AL096836; NID  
 A:Experimental source: strain Oresay  
 C:Genetics:  
 A:Gene: PAB1063

Query Match	7.2%	Score 143;	DB 2;	Length 409;
Best Local Similarity	20.7%	Pred. No. 0.0049;		
Matches	59;	Conservative	47;	Mismatches 83;
				Indels 96;
				Gaps 10
QY	11	AGEVVVHMSVLVILMLPTFMSLATM-LGQTGVGYPVAVVYMLLGAGVAMLLASLLAHELA	69	
DB	142	AGYLISL-----LYVTTLEBLNLGINKNTYLMNLAPSLG-----IISILGTHMG	186	
QY	70	HAVVARRAGVSESVT-----LMLFGYTAGLGEAKTPKKAFFRIARFGPTSL	117	
DB	187	HKIAA-----SHNVKSTPEYFIPPEPSFGLTGAVIRKSDIPTRNAEVDLGVSGPIAGL	241	
QY	118	ALSAFGALATLTAAGRTPAIV-----	139	
DB	242	LV-----ALPVTIIGLSKASVPIVNYLEKGTIYVGSLLPYGLMKVLQDLPNQVGIIL	286	
QY	140	--ISVAMWLATVNNLLGLFNNLLPGAPLDGGRLVRAVYLMRRHGSVVRGIGAGARARVVAL	197	
DB	297	HPRLAVGVG--VGIIATPFLNLLIPALQDGGHVARML-----PEKHARVLTY	340	

Thu Mar 23 10:41:48 2006

us-10-617-038-28.rpr

Page 7

```

Oy      198 VL--IALGLAEFVAGGLVGGMWLAIFGMWIFAAAREEETRISTQQ 240
          | : ||| | | : : | : | : | : : : :
Db      341 TLGLFTIGLAEFWPGWILMGILLILLMGRVGNPGALDEVSPLTTSR 385

```

Search completed: March 23, 2006, 05:26:35  
Job time : 16.6356 secs

This page Blank (uspto)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 04:56:43 ; Search time 100.515 Seconds  
(without alignments)  
2758.529 Million cell updates/sec

Title: US-10-617-038-28  
Perfect score: 1990  
Sequence: 1 MRD1PLGRINGFVVNVHMS.....QPEPTTSPQDADRFSFSDAG 393

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues  
Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_05.80:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1990	100.0	393	2 Q7Y73_MYCBO	Q7Y73 mycobacteri
2	1990	100.0	393	2 Q6187_MYCTU	Q6187 mycobacteri
3	497.5	25.0	409	2 Q8DJ3_SYNEI	Q8DJ3 synecococc
4	486.5	24.4	389	2 Q8YVX5_ANASP	Q8YVX5 anabaena sp
5	485	24.4	364	2 Q8PXA4_METMA	Q8PXA4 methanobac
6	485	24.4	364	2 Q8YU08_METAC	Q8YU08 methanobac
7	477	24.0	399	2 Q8YU03_ANASP	Q8YU03 anabaena sp
8	474.5	23.8	418	2 Q5UY6_HALMA	Q5UY6 haloarcula
9	471.5	23.7	381	2 Q5N0R6_SYNP6	Q5N0R6 synecococc
10	462.5	22.2	377	2 Q7NKH6_GLOVI	Q7NKH6 gloeobacter
11	459	23.1	390	2 Q9NHB6_HALSA	Q9NHB6 haloarcula
12	427	21.5	368	2 Q64D13_9ACCH	Q64D13 uncultured
13	425.5	21.4	379	1 Y528_SYNY3	Q5518 synecococcyt
14	424	21.3	422	2 Q7V6E0_PROMM	Q7V6E0 synecococcyt
15	418	21.0	362	2 Q29915_ARCFU	Q29915 archaeoglob
16	416	20.9	419	2 Q7U5H5_SYNPX	Q7U5H5 synecococc
17	407	20.5	368	2 Q64B18_9ACCH	Q64B18 uncultured
18	395.5	19.9	370	2 Q7J311_THET2	Q7J311 thermus the
19	391	19.6	370	2 Q6S1Q3_THET2	Q6S1Q3 thermus the
20	377.5	19.0	341	2 Q26907_MERTH	Q26907 methanobact
21	329	16.5	374	2 Q9R356_STRCO	Q9R356 streptomycet
22	328.5	16.5	339	1 Y392_METUA	Q57837 methanococc
23	328	16.5	372	2 Q92UP7_RHIME	Q92UP7 rhizobium m
24	326	16.4	328	2 Q6M020_METMP	Q6M020 methanococc
25	322.5	16.2	421	2 Q7VD44_PROMA	Q7VD44 prochloroc
26	313	15.7	356	2 Q5LSX8_SILPO	Q5LSX8 silicibacte
27	311	15.6	376	2 Q828J6_STRAM	Q828J6 streptomycet
28	299.5	15.1	289	2 Q5V077_HALMA	Q5V077 haloarcula
29	288	14.5	372	2 Q92283_RHIME	Q92283 rhizobium m
30	254	12.8	287	2 Q67SD6_SYWTH	Q67SD6 symbiodinace
31	224	11.3	408	2 Q7V2E2_PROMP	Q7V2E2 prochloroc

32	194.5	9.8	513	2 Q7N2W8_GLOVI	Q7N2W8 gloeobacter
33	187	9.4	266	2 Q7U14_RHOBA	Q7U14 rhodospirill
34	173	8.7	365	1 Y971_METUA	Q58381 methanococc
35	165.5	8.3	295	2 Q4NAJ8_PMICC	Q4NAJ8 archaeobacte
36	163.5	8.2	489	2 Q4NV60_9DELT	Q4NV60 anaeromyxob
37	160.5	8.1	368	2 Q8TUJ1_METAC	Q8TUJ1 methanobac
38	156.5	7.9	224	2 Q6N4M2_RHOPA	Q6N4M2 rhodospirill
39	155.5	7.8	607	2 Q8TK16_METAC	Q8TK16 methanobac
40	154.5	7.8	436	2 Q5JGJ8_PYRKO	Q5JGJ8 pyrococcus
41	154.5	7.7	222	2 Q4NPUS_9DELT	Q4NPUS anaeromyxob
42	151	7.6	203	2 Q6L2O8_PICPO	Q6L2O8 picophilus
43	150	7.5	360	2 Q7MEB6_VIBRY	Q7MEB6 vibrio vuln
44	149.5	7.5	284	2 Q97JML_CIOAB	Q97JML clostridium
45	149.5	7.5	365	2 Q63DTS_BACC2	Q63DTS bacillus ce

## ALIGNMENTS

RESULT 1					
ID	Q7Y73_MYCBO	PRELIMINARY;	PRT;	393 AA.	
AC	Q7Y73_				
DT	01-OCT-2003 (TREMBLrel. 25, Created)				
DT	01-OCT-2003 (TREMBLrel. 25, Last sequence update)				
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)				
DE	PROBABLE CONSERVED TRANSMEMBRANE ALANINE AND LEUCINE RICH PROTEIN.				
DE	OrderedLocustNames=Md2658c;				
GN	Mycobacterium bovis.				
OS	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium;				
OC	Mycobacterium tuberculosis complex.				
OX	NCBI_TaxID=1765;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=AF2122/97;				
RX	MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;				
RA	Garnier T., Sigelmeier K., Camus J.-C., Medina N., Manseur H.,				
RA	Pryor M., Durhoy S., Gondon S., Lacroix C., Monsenpe C., Simon S.,				
RA	Harris B., Aekin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,				
RA	Parthill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;				
RT	"The complete genome sequence of Mycobacterium bovis."				
RL	Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).				
DR	EMBL; BX248343; CAD94843.1 -; Genomic DNA.				
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.				
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.				
DR	InterPro; IPR000644; CBS.				
DR	InterPro; IPR008915; Peptidase_M50.				
DR	InterPro; IPR008915; Peptidase_M50.				
DR	Pfam; PF00571; CBS; 1.				
DR	Pfam; PF02163; Peptidase_M50; 1.				
DR	PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.				
KW	Complete proteome.				
SEQ	SEQUENCE 393 AA; 41480 MM; 6790E021215B6F98 CRC64;				
Query Match					
Best Local Similarity 100.0%; Pred. No. 6.5e-130; Length 393;					
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1 MRD1PLGRINGFVVNVHMSVLTLMFTWSLATMLPRTVGSGYDAVYVYMLGAGGAVMLL				60
DB	1 MRD1PLGRINGFVVNVHMSVLTLMFTWSLATMLPRTVGSGYDAVYVYMLGAGGAVMLL				60
QY	61 ASLAAHELAHVAVRRAGVSVESYTLWIFGCVTLAGEBAKTPKAAFRFAFGPATSLALS				120
DB	61 ASLAAHELAHVAVRRAGVSVESYTLWIFGCVTLAGEBAKTPKAAFRFAFGPATSLALS				120
QY	121 ATFGALATTLTAGVTPPAIVISVAMWLTAVNLGLFNLPLGAPLDGGRLVAVYLMRRHGD				180
DB	121 ATFGALATTLTAGVTPPAIVISVAMWLTAVNLGLFNLPLGAPLDGGRLVAVYLMRRHGD				180
QY	181 SVRAGIGARRAGVVALVTLNLGLAEFVAGLVGVWMLAFIGWPIFAAAREESTRISTQQ				240



Db 181 SVRAGIGAAAGRVVVALVLIAGLAEFVAGGLVGWVLAFIGWFIFAAREETRISTQQ 240  
Qy 241 LFAGRVADAMTAOPHTAPGWINVEDFIORVYLGERSAIPVADRDGSIIGLVLRQLRD 300  
Db 241 LFAGRVADAMTAOPHTAPGWINVEDFIORVYLGERSAIPVADRDGSIIGLVLRQLRD 300  
Qy 301 VAPSRSTTSVGDIALPLHSVPTARPOEPLTALLERMAPLGRSRALVTBGSAAVGIWTP 360  
Db 301 VAPSRSTTSVGDIALPLHSVPTARPOEPLTALLERMAPLGRSRALVTBGSAAVGIWTP 360  
Qy 361 SDVARLIDVYRLAQPEPTFTTSPQADRFSDAG 393  
Db 361 SDVARLIDVYRLAQPEPTFTTSPQADRFSDAG 393

RESULT 2  
ID 006187 MYCTU PRELIMINARY; PRT; 393 AA.  
AC 006187; Q7D6V5;  
DT 01-JUL-1997 (TReMBLrel. 04, Created)  
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)  
DT 13-SEP-2005 (TReMBLrel. 31, Last annotation update)  
DE PROBABLE CONSERVED TRANSMEMBRANE ALANINE AND LEUCINE RICH PROTEIN  
(Hypothetical protein).  
GN OrderedLocNames=MT2700, RV2625C;  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;  
OC Mycobacterium tuberculosis complex.  
OX NCBI\_TaxID=1773;  
RN (1)  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=H37RV;  
RX MEDLINE=96295887; PubMed=9634230; DOI=10.1038/31159;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C.M.,  
RA Harris D.E., Gordon S.V., Eigmler K., Gae S., Barry C.E. III,  
RA Tekela F., Badcock K., Basham D., Brown D., Chillingworth T.,  
RA Connor R., Davies R.M., Devlin K., Feltwell T., Gentles S., Hamlin N.,  
RA Holroyd S., Hornby T., Jagsels K., Krogh A., McLean J., Moule S.,  
RA Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,  
RA Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;  
RA "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544 (1998).  
RN (2)  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=CDC 1551 / Oshkosh;  
RX MEDLINE=22206494; PubMed=12218036;  
RX DOI=10.1128/JB.184.19.5479-5490.2002;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J.D., DeBoy R.T., Dodson R.J., Gwinn M.L., Haft D.H.,  
RA Hickey E.K., Kolonay R.T., Nelson W.C., Umayam L.A., Esmolaeva M.D.,  
RA Salzberg S.L., Delcher A., Uitterback T.R., Weidman J.F., Khouri H.M.,  
RA Gill J., Minkina A., Bishai W., Jacobs W.R. Jr., Venter J.C.,  
RA Fraser C.M.;  
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL J. Bacteriol. 184:5479-5490 (2002).  
DR EMBL, BX842580; CAB0617.1; -; Genomic\_DNA.  
DR EMBL, AE000516; AAK47016.1; -; Genomic\_DNA.  
DR PIR, H70572; H70572.  
DR TIGR, MT2700; -;  
DR TUBERCULATC, RV2625C; -;  
DR GO, GO:0004222; F:metalloendopeptidase activity; IEA.  
DR GO, GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro, IPR000644; CBS.  
DR InterPro, IPR008915; Peptidase\_M50.  
DR InterPro, IPR006025; Pept\_M\_zn\_BS.  
DR Pfam, PF00571; CBS; 1.  
DR Pfam, PF02163; Peptidase\_M50; 1.  
DR SMART, SM00116; CBS; 1.  
DR SMART, SM00116; CBS; 1.

DR PROSITE, PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 393 AA; 41480 MW; 6790E021215B6F98 CRC64;  
Query Match 100.0%; Score 1990; DB 2; Length 393;  
Best local similarity 100.0%; Pred. No. 6.5e-130; Indels 0; Gaps 0;  
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MRDAIPGRIAGFVNVNWSVTLVLMFTWSLATMLPGTVGGYDAVYVYMLGAGAVWLL 60  
Db 1 MRDAIPGRIAGFVNVNWSVTLVLMFTWSLATMLPGTVGGYDAVYVYMLGAGAVWLL 60  
Qy 61 ASLIAHEIAHAHVARRAGVSYESYTLMLFGCVTNLAGSEAKTPKAFRIAPGPTSLALS 120  
Db 61 ASLIAHEIAHAHVARRAGVSYESYTLMLFGCVTNLAGSEAKTPKAFRIAPGPTSLALS 120  
Qy 121 ATFGALAITTAGVTPATIVISVAMWLTATVNLGLFNLPGAPLDGRLVRAVYLMRRHGD 180  
Db 121 ATFGALAITTAGVTPATIVISVAMWLTATVNLGLFNLPGAPLDGRLVRAVYLMRRHGD 180  
Qy 181 SVRAGIGAAAGRVVVALVLIAGLAEFVAGGLVGWVLAFIGWFIFAAREETRISTQQ 240  
Db 181 SVRAGIGAAAGRVVVALVLIAGLAEFVAGGLVGWVLAFIGWFIFAAREETRISTQQ 240  
Qy 241 LFAGRVADAMTAOPHTAPGWINVEDFIORVYLGERSAIPVADRDGSIIGLVLRQLRD 300  
Db 241 LFAGRVADAMTAOPHTAPGWINVEDFIORVYLGERSAIPVADRDGSIIGLVLRQLRD 300  
Qy 301 VAPSRSTTSVGDIALPLHSVPTARPOEPLTALLERMAPLGRSRALVTBGSAAVGIWTP 360  
Db 301 VAPSRSTTSVGDIALPLHSVPTARPOEPLTALLERMAPLGRSRALVTBGSAAVGIWTP 360  
Qy 361 SDVARLIDVYRLAQPEPTFTTSPQADRFSDAG 393  
Db 361 SDVARLIDVYRLAQPEPTFTTSPQADRFSDAG 393

RESULT 3  
ID 00DJC3 SYNEL PRELIMINARY; PRT; 409 AA.  
AC 00DJC3  
DT 01-MAR-2003 (TReMBLrel. 23, Created)  
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
DE T111304 protein.  
GN OrderedLocNames=t111304;  
OS Synechococcus elongatus (Thermosynechococcus elongatus).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
OX NCBI\_TaxID=32046;  
RN (1)  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=BP-1;  
RX MEDLINE=2225144; PubMed=12240834;  
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
RA Watanabe A., Iriyuchi M., Kawashima K., Kimura T., Kashiida Y.,  
RA Kiyokawa C., Kohara M., Matsunoto M., Matsuno A., Nakazaki N.,  
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
RA "Complete genome structure of the thermophilic cyanobacterium  
RT Thermosynechococcus elongatus BP-1.";  
RL DNA Res. 9:123-130 (2002).  
DR EMBL, BA000039; BAC08856.1; -; Genomic\_DNA.  
DR GO, GO:0004222; F:metalloendopeptidase activity; IEA.  
DR GO, GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro, IPR000644; CBS.  
DR InterPro, IPR008915; Peptidase\_M50.  
DR InterPro, IPR006025; Pept\_M\_zn\_BS.  
DR Pfam, PF00571; CBS; 1.  
DR Pfam, PF02163; Peptidase\_M50; 1.  
DR SMART, SM00116; CBS; 1.  
DR PROSITE, PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
KW Complete proteome.  
SQ SEQUENCE 409 AA; 45373 MW; 7C192803184EDA68 CRC64;

QY	1	MRDAIPGRIAGFVNVNHSVLTLMFTSLTMTLPGTVGVY-----AVYVWMLGAG	54
Db	17	MNGTIRAGNIFGIFPYIHPSPFLVLGLVTSYS---GGLAEFPOLSGRLAVY---LGLA	70
QY	55	GAWVLLSLAHSLAHVVAVRBAGVSVESTLMFGVTLAGEAKTPKAFAFLAFAGPA	114
Db	71	TALLLPASVVAHEGHFSFVAIRQDINASTLTLFPGSLASEBSKTPAGAFVVALNGPL	130
QY	115	TSLSLSTFGALATTLTGLVTRTPAIVISAVMWLTAVNLLGLFNLLPGAPLDGRLVAYL	174
Db	131	VSLLLCGVITVTIGTTAVTGPRLAIIIGV---LASVNLALPFLNIPGLPDGNNILKAIY	187
QY	175	WRHGDVSRAGIGAAPRGVVALVLLALGLAEFVAGLVGVGWLAFGWTFFAAAREET	234
Db	188	WKVTGNPKYGTFSASRGVIGFWAIVLSGIFPIILYFGSSFANVNNLLIGFFLLNAGNMAQ	247
QY	235	RISTQQLFAGRVADNATQAPHTAPRGAINVEDFIQRVVL-GERHSAFVADRGSTIGLV	293
Db	248	FARVQELKTGLTADAVATVTSPIVSAHLSIRFDDOIIQGWRRFLVNNAGOLGAI	307
QY	294	ALROLRDVAPRSRSTTSVGDIALPLHSVPARPPEPTALLERMAPLSPRSALVTGESA	353
Db	308	ALDDLKRNIFTSWETQIQQVMRIQST-ITKSSQPLLEVQQLLEQKLSALPVIIDNGV	366
QY	354	VGVITPESDVARLLIDVYRLAQPPEP	377
Db	367	LLGILIERKAITQLDQ--NGTQPNP	388
RESULT 5			
ID	Q8PX44	METWA PRELIMINARY; PRT; 364 AA.	
AC	Q8PX44		
DT	01-OCT-2002	(TReMBLrel. 22, Created)	
DT	01-OCT-2002	(TReMBLrel. 22, Last sequence update)	
DT	01-MAR-2004	(TReMBLrel. 26, Last annotation update)	
DN		Conserved protein.	
GN		Ordered locus names=MM1318;	
OC		Methanosarcina mazel (Methanosarcina fistia).	
OC		Archaea; Buryarchaeota; Methanomicrobis; Methanosarcinales;	
OC		Methanosarcinaceae; Methanosarcina.	
OX		NCBI_TaxID=2209;	
RN		[1]	
RP		NCUIEOTIDE SEQUENCE.	
RC		STRAIN=Goel / G01 / ATCC BAA-199 / DSM 3647 / OCM 88;	
RX		MEDLINE=22120827; PubMed=12125824;	
RA		Deppe-meier U., Johann A., Hartsoch T., Merkl R., Schmitz R.A.,	
RA		Martinez-Artis R., Henne A., Wiese A., Beunier S., Jacobl C.,	
RA		Brueggemann H., Lienard T., Christmann A., Boemeke M., Steckel S.,	
RA		Bhattacharyya A., Lykdis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,	
RA		Fritz H.-J., Gottschalk G.;	
RT		"The genome of Methanosarcina mazel: evidence for lateral gene	
RT		transfer between Bacteria and Archaea.";	
RL		J. Mol. Microbiol. Biotechnol. 4:453-461(2002).	
DR		EMBL; AE013364; AAM31014.1; -; Genomic DNA.	
DR		GO: GO:0004222; F:metalloendopeptidase activity; IEA.	
DR		GO: GO:0006508; P:proteolysis and peptidolysis; IEA.	
DR		InterPro; IPR000644; CBS.	
DR		InterPro; IPR008915; Peptidase_M50.	
DR		InterPro; IPR006025; Pept_M_Zn_BS.	
DR		Pfam; PF00571; CBS; 1.	
DR		Pfam; PF02163; Peptidase_M50; 1.	
DR		SMART; SM00116; CBS; 2.	
DR		PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.	
KW		Complete proteome.	
SO		SEQUENCE 364 AA; 40168 MW; 80E1ADA5173A9CD9 CRC64;	
Query Match 24.4%; Score 485; DB 2; Length 364;			
Best Local Similarity 31.5%; Pred. No. 1,5e-25;			
Matches 121; Conservative 85; Mismatches 136; Indels 42; Gaps 12;			
QY	1	MRDAIPGRIAGFVNVNHSVLTLMFTSLTMTLPGTVGVY-----PAVYVWMLG	52

Qy	Db	Sequence	Score	DB 2	Length			
Qy	Db	1 MMSKLTGSLGAIPIYKLTITFLILFVPIYFA-----INPYPGQEVPEPSTKXMS 54	24.4%	Score 4851	DB 2; Length 364;			
Qy	53	AGCAVMLLASLAAHELAHVAVRRAGVSVEYVLTMLFGVTAALGCEAKTPKCAFRIFAAG 112						
Db	55	SITATILFASILHLEHAHSLYLAQYGVNIESITLFLFGGVSSMNEELPRDGCBAKMAFAG 114						
Qy	113	PATSLALSATFGALATITLGA---VTRPAI---VISAVMWLATVNLILGLFNLLPGAPLD 165						
Db	115	PFTSLIVI---GSICLLLYRYIIISNPALSENPEVYLTITLGMNVLGI FNLLPAFPMD 170						
Qy	166	GGRLVRAVIMRRHGSVVRAGIGAAARGRVAVALVILGLAEFPAAGLNGVWMLAFGMFI 225						
Db	171	GGRVLRSPFAARM-SYKATKSKAAVAKKFPALIMAFGI-----LIGLWFPPLALFI 222						
Qy	226	FAAAREEETRISTQOLFAGVRVADAMTAOPHTAPGMINVEDFIQRYVLGERHSAYEPVADR 285						
Db	223	YAGABEEERSSTQASVSLNIIVKIDITKXVSVSPSPMNVEDLIQ-FMEFKKHGYVVE- 280						
Qy	286	DGSITGLVALROLRDVAPSRSTTSVGI-ALPLHSVPT-ARDPEPLTALLEMAPLGR 343						
Db	281	SGNLKGIIVPTDIOGVPTIDRPVRRVSDIMTRDIIISVPSDAQSDVLKLVTSKNI----- 345						
Qy	344	SGALVTGSAVVGIVTPSPDVALI 367						
Db	336	GRVLVINGSLVGLSRTDLVRL 359						
RESULT 6								
OSTUD08 METAC								
ID	OSTUD08_METAC	PRELIMINARY;	PRT;	364	AA.			
AC	OSTUD08;							
DT	01-JUN-2002	(Tremblrel, 21, Created)						
DT	01-JUN-2002	(Tremblrel, 21, Last sequence update)						
DT	01-MAR-2004	(Tremblrel, 26, Last annotation update)						
DE	Hypothetical protein MA0007.							
GN	Ordered locus names=MA0007.							
OS	Methanosarcina acetivorans.							
OC	Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;							
OC	Methanosarcinaceae; Methanosarcina.							
CC	NCBI_TaxID=2214;							
RN	[1]							
RP	NUCLEOTIDE SEQUENCE.							
RC	STRAIN=C2A / ATCC 35395 / DSM 2834;							
RX	MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;							
RA	Galagan J.E., Nuebaum C., Roy A., Endrizzi M.G., MacDonald P.,							
RA	Fitzhugh W., Calvo S., Engels R., Smitnov S., Atencio D., Brown A.,							
RA	Allen N., Naylor J., Strange-Thomann N., Deatellano K., Johnson R.,							
RA	Linton L., Mcwan P., McEernan K., Talamas J., Titrill A., Ye W.,							
RA	Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,							
RA	Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,							
RA	Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,							
RA	Springer T.A., Unayam L.A., White O., White R.H., de Macario E.C.,							
RA	Ferry J.G., Darrell K.F., Jing H., Macario A.J.U., Paulsen I.T.,							
RA	Pritchett W., Sowers K.R., Swanson R.V., Zinder S.R., Lander E.,							
RA	Metcalfe M., Birren B.;							
RT	"The genome of Methanosarcina acetivorans reveals extensive metabolic							
RT	and physiological diversity."							
RL	Genome Res. 12:533-542(2002).							
DR	EMBL, AB010657; AAM03461.1; -, Genomic DNA.							
DR	GO; GO:0004222; F:metalloendopeptidase activity; IEA.							
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.							
DR	InterPro; IPR008915; Peptidase M50.							
DR	InterPro; IPR006025; Pept_M_Zn_BS.							
DR	Pfam; PF00571; CBS; 1.							
DR	Pfam; PF02163; Peptidase_M50; 1.							
DR	SMART; SM00116; CBS; 2.							
DR	PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.							
DR	Complete proteome.							
SC	SEQUENCE 364 AA; 39671 MW							

[illegible]

	RESULT 7		
Q8YOD3	_ANASP	PRT;	399 AA.
ID	Q8YOD3_ANASP PRELIMINARY;		
AC	Q8YOD3;		
DT	01-MAR-2002 (TREMBLrel. 20, Created)		
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)		
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)		
DE	A113900 protein.		
GN	OrderedLocustNames=alrj3900;		
OS	Anabaena sp. (strain PCC 7120).		
OC	Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.		
NCBI_TaxID	=103690;		
RN	[1]		
RL	NUCLEOTIDE SEQUENCE.		
RX	MEDLINE=21595285; PubMed=11759840;		
RA	Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,		
RA	Weetababe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,		
RA	Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,		
RA	Nakazaki N., Shimojo S., Sugimoto M., Takazawa M., Yamada M.,		
RA	Yasuda M., Tabata S.;		
RT	"Complete genomic sequence of the filamentous nitrogen-fixing		
RT	Cyanobacterium anabaena sp. strain PCC 7120."		
RL	DNA Res. 8:205-213(2001).		
EMBL:	BA000019; BAB75599.1; -; Genomic_DNA.		
DR	PIR: AE2293; AE2293.		
DR	GO: GO:0004222; F:metalloendopeptidase activity; IEA.		
DR	GO: GO:0006508; P:proteolysis and peptidolysis; IEA.		
DR	InterPro: IPR000644; CBS.		
DR	InterPro: IPR008915; Peptidase_M50.		
DR	Pfam: PF00571; CBS; 1.		
DR	Pfam: PF02163; Peptidase_M50; 1.		
DR	SMART: SM00116; CBS; 1.		
DR	PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.		
KW	Complete proteome.		
SQ	SEQUENCE 399 AA; 43105 MW; EC3C4E458F8P200D CRC64;		
Query Match	24.0%; Score 477; DB 2; Length 399;		
Best Local Similarity	34.1%; Pred.No. 6.le-25;		
Matches 120; Conservative 59; Mismatches 163; Indels 10; Gaps 4;			

```

QY 25 LMLFTWLSATMLPGTV-GGYPAVVYWLIGAGGAVMLLSLAHELAHAVARAGVSSES 83
DB 21 LMFVLLGLATLNFVAVYQEMGTVTAVTAGLIMSLFFGSLVLLHGHSLAARSQGIKXNS 80
QY 84 VTLMLPGVTLALGGEAKTPKAAFRAPAPATSLSLSTFGALATTLAAGVRPAIVISA 143
DB 81 ITLPLFGGIALAEESKTPGKAFQVAILAPLVISGI--FLLRLRGSTVSDSPSVMV 137
QY 144 WMLATVNLGLFENLLPGALPDGRLVRAVYLWRBHDGVSRAIGAAARAGVALVLIAG 203
DB 138 GDLAINLVALLFNLLPGPLDGGVVKALMLQITGDRQAYHMAKAGQIIGYALNAG 197
QY 204 LA-EVAVAGLVGVWLAFIIGWFIFAAREETRISTQQLFAGVRVADAMTAQPHAPGI 262
DB 198 FAVDEFTRELVTGLMTALLGFGVNRANSYDRVTTLQETLLREKAVDAMTRDFRVIDANQ 257
QY 263 NVEDTQRYVGERRSAYVADRDGSIITGLVALROLRDVAPRRSTSVGDIATLPLHSP 322
DB 258 TLRSFADSYLATSPEVYFAASDGRYRGMVAIEDRLVERSEWETQTHSLAHPLEIP 317
QY 323 TARPQEPITLALERMARPLGPRSRALVTGSAVVGIVTPSDVARLIDVYRLAQ 374
DB 318 TYAESTALAGVINKLENEQLPRVTVLTPAGAVAGIIRDGI-----VGLALAQ 364

```

## RESULT 8

OSUYV6 HALMA PRELIMINARY; PRT; 418 AA.

```

AC OSUYV6;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=ernAC2755;
OS Haloarcula marismortui (Halo bacterium marismortui).
OC Halobacteriales; Euryarchaeota; Halobacteriia; Halobacteriaceae; Haloarcula.
OX NCBI_TaxID=2238;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 43049;
RX PubMed=15520287; DOI=10.1101/gr.2700304;
RA Baliga N.S., Bonneau R., Facciotti M.T., Pan M., Glusman G.,
RA Deutsch E.W., Shannon P., Chiu Y., Meng R.S., Gan R.R., Hung P.,
RA Date S.V., Marcotte E., Hood L., Ng W.V.;
RT "Genome sequence of Haloarcula marismortui: a halophilic archaeon from
RT the Dead Sea."
RL Genome Res. 14:2221-2234(2004).
DR EMBL; AY596297; AAV47518.1; -; Genomic DNA.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000644; CBS.
DR InterPro; IPR008915; Peptidase_M50.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF00571; CBS; 1.
DR Pfam; PF02163; Peptidase_M50; 1.
DR SMART; SM00116; CBS; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR Complete proteome; Hypothetical protein.
SQ SEQUENCE 418 AA; 4515 MW; 2C6256D1E2C227AD CRC64;

```

Query Match 23.8%; Score 474.5; DB 2; Length 418;

Best Local Similarity 33.7%; Pred. No. 9.5e-25;

Matches 137; Conservative 65; Mismatches 160; Indels 45; Gaps 16;

```

QY 7 LGRIAGFVNVAVSVLILMLFTWLSAT-----MLPGTVG-----YPAVVYWLIG 52
DB 23 ICSAGIPIQLDLTFLVLPFLPAMIIGTQIEQTTELTNGTLNAGLDVAVLTGALVWVLG 82
QY 53 AGGAVMLASLALHELAHAVARAGVSVESTLMLPGVTLALGGEAKTPKAAFRIPAPG 112
DB 83 IGAALGLFTGVVLAHEIGHSLVAIRYGFPIIDSLITMLFGIADLSEMPEDMKDELVIATAG 142

```

```

QY 113 PATSLALSAT-FGALATTLAAGVRPAIVIS--VAMMLATVNLGLFENLLPGALPDGCR 168
DB 143 PIVSIAIGAVCFVAFQIIPSGAGT--IVESTRFIIQYDALNMLAAEFNMLPGFPMDCGR 200
QY 169 LVRAVLMRHDGVSRAIGAAARAGRVVLIAGLAEFVAGLVGVWLAFIIGFIPA 228
DB 201 VLRALLARR-SYATATTIAEVGCIFAVFIQGLFQI--FV---LGNFLLGLAFIYIG 253
QY 229 AREETRISTQQLFAGVRVADAMTAQPH--TAPGMINVEDTQRYVGERRSAYVADRD 286
DB 254 AAGBSRQTSMAAPFGVAVADVMTPADVHTTVADMSVRELQI-MFRERHTGYV-V-RKS 311
QY 287 GSITGLVALROLRDVAPRRSTSVGDIATLPLHSPVTPARPOEPLTALERMARPLGPR-- 344
DB 312 GEVGLVLTLEDAARAQVEVERAAYTGVDWMT--EITITISPE---TDVMDALTSIQNSVG 366
QY 345 RALVT-EGSAVVGIVTPSDVARLIDV-----YRLAQEPFTTSPQ 384
DB 367 RLVTDEDSFEGELTRSDIMTALSTIKSSDYTAIGSEETETVRP 413

```

## RESULT 9

QSNOR6 SYNPE PRELIMINARY; PRT; 381 AA.

```

AC QSNOR6;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=BYC1914.d;
OS Synechococcus sp. (strain PCC 6301) (Anacystis nidulans).
OC Bacteria; Cyanobacteriia; Chroococcales; Synechococcus.
OX NCBI_TaxID=269084;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PCC6301;
RA Sugita M.;
RT "Complete genome structure of the unicellular cyanobacterium Anacystis
RT nidulans 6301 (Synechococcus sp. PCC6301).";
RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP008223; BAD80104.1; -; Genomic DNA.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000644; CBS.
DR InterPro; IPR008915; Peptidase_M50.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF00571; CBS; 1.
DR Pfam; PF02163; Peptidase_M50; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR Complete proteome; Hypothetical protein.
SQ SEQUENCE 381 AA; 41197 MW; 7341790A36BCECF CRC64;

```

Query Match 23.7%; Score 471.5; DB 2; Length 381;

Best Local Similarity 32.0%; Pred. No. 1.4e-24;

Matches 113; Conservative 66; Mismatches 131; Indels 43; Gaps 7;

```

QY 29 TWSLATMLPGTVGVPAVVYWLIGAGGAVMLLSLAHELAHAVARAGVSVESTLML 88
DB 20 TWS-----FGLMGALVW-----ALLPASYLAHELGHSLIARAGIRVSSITFL 66
QY 89 FGVYTLALGGEAKTPKAAFRIPAPATSLASATFGALATTLAAGVRPAIVISVAMW--- 145
DB 67 FGVVAALIREERTPGAFWVAIAGPLVSFALAL-----LLISQLMWDAG 111
QY 146 -----LATVNLGLFENLLPGALPDGRLVRAVYLWRBHDGVSRAIGAAARAGRYVA 196
DB 112 SPAQVLSLNLGRSLSTIAVFNLLPGPLDGGVVKALMLKMTGDRYRVAHMAANSGRLLS 171
QY 197 LVLTALGL-AEFVAGLVGVWLAFIIGWFIFAAREETRISTQQLFAGVRVADAMTAQPH 255
DB 172 AIAVAIGFSPFLDGGSGVWLALHGFGRNATAYVPTTTLQOAILAIGASEMSRRY 231
QY 256 HTAPGMINVEDTQRYVGERRS-AYPVADRDGSIITGLVALROLRDVAPRRSTTSVGI 314

```

Dd		232	RVLBSLTLRQFAELLITTEDEGFAFYVAS-DGRYKGI SLATLTQTERSQDRLLTLDL	290
Oy		315	ALPLHVSPTARPOEPLTALLERMAFLGPRSRALTVEGSAAWGIVTPSDVARLI	367
Dd		291	AEPFRRLALRETNLAQAIALAQTAFQSYYTVLVLPSSAGVAGI IDHAVIDQL	343
		RESULT 10		
		07NKH6 GLOVI		
		07NKH6 GLOVI PRELIMINARY;	PRT,	377 AA.
Dt		01-MAR-2004 (TREMBLrel. 26, Created)		
Dt		01-MAR-2004 (TREMBLrel. 26, Last sequence update)		
Dt		01-MAR-2004 (TREMBLrel. 26, Last annotation update)		
Gn		Glrl502 protein.		
Dn		OrderedLocusNames=glrl1502;		
OC		Gloeobacter violaceus.		
OC		Bacteria; Cyanobacteria; Gloeobacteriales; Gloeobacter.		
OX		NCBI_TaxID=33072;		
RP		NUCLEOTIDE SEQUENCE [large scale genomic DNA].		
RC		STRAIN=PCC 7421.		
RX		MEDLINE=P2977040; PubMed=14621292;		
RA		Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyashita H., Tsuchiya T.,		
RA		Saemoto S., Watanabe A., Kawashima K., Kishida Y., Kiyokawa C.,		
RA		Kohara M., Matsunoto M., Matsuno A., Nakazaki N., Shimo S.,		
RA		Takeuchi C., Yamada M., Tabata S.;		
RT		"Complete genome structure of Gloeobacter violaceus PCC 7421, a		
RT		cyanobacterium that lacks thylakoids."		
RL		DNA Ref. 10:137-145(2003).		
DR		EMBL; BA000045; BAC89443.1; -; Genomic_DNA.		
DR		GO; GO:0004222; F:metalloendopeptidase activity; IEA.		
DR		GO; GO:0006508; P:proteolysis and peptidolysis; IEA.		
DR		InterPro; IPR00644; CBS.		
DR		InterPro; IPR008915; Peptidase_M50.		
DR		InterPro; IPR006025; Pept_M_zn_BS.		
DR		Pfam; PF02163; Peptidase_M50; 1.		
DR		Pfam; PS00142; ZINC_PROTEASE; UNKNOWN_1.		
SW		Complete proteome.		
SO		SEQUENCE 377 AA; 40561 MW; F37B3BBDDBEEBBF CRC64;		
		Query Match	23.2%; Score 462.5; DB 2; Length 377;	
		Best Local Similarity	33.4%; Pred. No. 5, 8e-24;	
		Matches 129; Conservative 63; Mismatches 163; Indels 31; Gaps 13;		
Oy		2	RDALPLGLIAG--PVVNVHMSVLVILMLFTWSLAMLEGTGVGYPAVVTWLLGAGAVML	59
Dd		8	KGAIRAGSLFGIPFPIIDVSM--FLIIAPFTWYSYGALGNHDWGTIVFWLPGFVSALL	65
Oy		60	LASILAHEIAHVAVVRRAVSVESVTLMFGSVTLALGGCAKPKKAFLRIAPATSIAL	119
Dd		66	FASVILHELHGSPAIAGCIRVQSLSLTFGVADIREERSRPMKALVALLAGLVSIAL	125
Oy		120	SATFGAL--ATTLAGVTRPA--IVISAVMALTAVNLIGLFNLLPGAPLDGELVAYIMR	176
Dd		126	FGLFYGLEQALSIAG---PAGAVVEL---LAVNALAILFNMLPGILDGGVTLKALVWG	179
Oy		177	RHGDSVPARGICGARAGRVAVALVILALGLAEFPAGLVGCVWLAFGWFLPAAAREEFRI	236
Dd		180	VTGNOYKIRIFAGIGTGGOGVGLMMGLG--FYIGNF-NGLMPAIIIGWFVFSNARYSEYA	236
Oy		237	STQQLFAVNRVADAMTAQPHTPAPGINVEDFIQRVLERHSAPVADBDGSLTGLV---	293
Dd		237	RVQGSLSTLQAQVAAVRTEPAVPAHASLSKSFADLYALSQQVSLTYDIDGVLVERIDRG	296
Oy		294	-ALRQLRVASRSRSTTSVDITALPLHSVPTARPOEPLTALLERMAPLG-PPSRALVTEG	351
Dd		297	ALLRRPTET----WAITPSAWMGVVDASETYAADQRPDLDEVLTRIQEKRLPRIPIVLQPSG	352
Oy		352	SAVGIVTPSDVARLIDVYRLAQPDP 377	

[illegible]

```

Db      348 LVVDADGEMVGLTRSD---LMDAFGIQ 373

RESULT 12
ID      Q64D13_9ARCH PRELIMINARY; PRT; 368 AA.
AC      Q64D13;
DT      25-OCT-2004 (TREMBLrel. 28, Created)
DT      25-OCT-2004 (TREMBLrel. 28, Last sequence update)
DT      25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE      Zn-dependent proteases.
GN      ORFNames=GZ18C8_21;
OS      uncultured archaeon GZfos18C8.
OC      Archaea; environmental samples.
OX      NCBI_TaxID=285411;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RX      PubMed=15353801; DOI=10.1126/science.1100025;
RA      Hallam S.D., Putnam N., Preston C.M., Dettler J.C., Rokhsar D.,
RA      Richardson P.M., Delong E.F.;
RT      "Reverse methanogenesis: testing the hypothesis with environmental
RT      genomes.";
RL      Science 305:1457-1462 (2004).
RN      [2]
RP      NUCLEOTIDE SEQUENCE.
RA      Putnam N., Dettler J.C., Richardson P.M., Rokhsar D.;
RL      Submitted (AUG-2004) to the EMBL/Genbank/DBJ databases.
DR      EMBL; AY714825; AU082544.1; -, Genomic_DNA.
DR      GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR      GO; GO:0008233; F:peptidase activity; IEA.
DR      GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR      InterPro; IPR00644; CBS.
DR      InterPro; IPR008915; Peptidase_M50.
DR      InterPro; IPR006025; Pept_M_Zn_BS.
DR      Pfam; PF00571; CBS; 1.
DR      Pfam; PF02163; Peptidase_M50; 1.
DR      SMART; SM00116; CBS; 2.
DR      PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW      Protease.
SQ      SEQUENCE 368 AA; 40972 MW; 76ED9972F073FC CRC64;

Query Match 21.5%; Score 427; DB 2; Length 368;
Best Local Similarity 29.8%; Pred. No. 1.6e-21;
Matches 114; Conservative 76; Mismatches 165; Indels 28; Gaps 10;

Qy      1 MEDAIPLRIGFVNVNMSVLIIMLFWSLATMLPGVGYPAV---VTMLGAGGA 56
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1 M KTSIQIKVWGIPKIHISFLVLPVVFAN--DAVFQAVASVLTAEVATLT 57
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      57 VMLASLAAHELAAVAVRAGVSVESTLMFGVTLAGEAKTPKAFRIAPAPATS 116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      58 ILLSVCVLLHSGHGWAVRAGVIGIRSTITLLIGIAAEEVPRPRKEMRISIGPLVS 117
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      117 LALS-----ATFGALATITLAVGRTPAIVSVAMWLTATVLLGLFNLLPGAPLDGRVLR 171
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      118 LTIGVLCVLAIVGLGIYDMVQSTISHSFLMSIAYINIVAPFNLLIPAFPMDSGRVLR 177
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      172 AYLRRHGDSTVAGIGAAPARVVALVLIAGLAEFVAGLVGWLAIFGFIIPAARE 231
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      178 AW-YAGHPYLAATKAAVHIGMFAIVMGVL-----GLFASITLIIAIFFIYIGASE 228
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      232 EETRISTQQLFAGVAVADAMTAQPHAPGMINVEDFIQRYVYGERHSAPVADR-DGSIT 290
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      229 EEKTYEVSTLEGIRNDRMTREIAIVQDNLTISELL-RLMEKKHLGIPVVDQFTGKI 287
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      291 GLVALRQLADVAPSRSTSTVGDIAPLHSVPTAPQPEELTALRMAPLGRSPALYTE 350
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      288 GIVTFDIDRSVPMSEHGVLVREV-MAKVVFIPEADAMAL--KMSSTENVGQLVQD 344
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      351 GSAVVGIVTSPVAKLIDV--YR 371
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      345 RGAITGIIVSRDITLRSIEVLGYR 367
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 13
ID      Y528 SYN3 STANDARD; PRT; 379 AA.
AC      O5518;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      13-SEP-2005 (Rel. 48, Last annotation update)
DE      Hypothetical protein gl10528 (EC 3.4.24.-).
GN      OrderedLocustNames=gl10528;
OS      Synecocystis sp. (strain PCC 6803).
OC      Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX      NCBI_TaxID=1148;
RN      [1]
RP      NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX      MEDLINE=96127529; PubMed=8590279;
RA      Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA      Sugitara M., Tabata S.;
RT      "Sequence analysis of the genome of the unicellular cyanobacterium
RT      Synecocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT      region from map positions 64k to 92k of the genome.";
RL      DNA Res. 2:153-166 (1995).
RN      [2]
RP      NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX      MEDLINE=97061201; PubMed=8905231;
RA      Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA      Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,
RA      Hoshino T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA      Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA      Tabata S.;
RT      "Sequence analysis of the genome of the unicellular cyanobacterium
RT      Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT      entire genome and assignment of potential protein-coding regions.";
RL      DNA Res. 3:109-136 (1996).
CC      -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC      -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC      -1- SIMILARITY: Belongs to the peptidase M50B family.
CC      -1- SIMILARITY: Contains 2 CBS domains.
CC      -----
CC      This Swiss-Prot entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use as long as its content is in no way modified and this statement is not
CC      removed.
CC      -----
DR      EMBL; BA000022; BAA10876.1; -, Genomic_DNA.
DR      PIR; S76029; S76029.
DR      InterPro; IPR000644; CBS.
DR      InterPro; IPR006025; Pept_M_Zn_BS.
DR      InterPro; IPR008915; Peptidase_M50.
DR      Pfam; PF00571; CBS; 1.
DR      Pfam; PF02163; Peptidase_M50; 1.
DR      SMART; SM00116; CBS; 2.
DR      PROSITE; PS00142; ZINC_PROTEASE; 1.
KW      CBS domain; Complete proteome; Hydrolase; Hypothetical protein;
KW      Metal-binding; Metalloprotease; Protease; Repeat; Transmembrane; zinc.
FT      TRANSMEM 20
FT      TRANSMEM 40
FT      TRANSMEM 54
FT      TRANSMEM 115
FT      TRANSMEM 148
FT      TRANSMEM 212
FT      TRANSMEM 260
FT      DOMAIN 310
FT      DOMAIN 327
FT      DOMAIN 375
FT      ACT SITE 76
FT      METAL 75
FT      METAL 79
FT      METAL 79
SQ      SEQUENCE 379 AA; 40465 MW; 1CC3251650078ECB CRC64;

Query Match 21.4%; Score 425.5; DB 1; Length 379;
Best Local Similarity 30.0%; Pred. No. 2.2e-21;
Matches 115; Conservative 70; Mismatches 147; Indels 51; Gaps 11;

```



QY 1 MRDAIPGIRING--FVNVNHSVLTITLFTMSLATMNL---PGTVGYPVAVVYLLGAG 55  
DB 11 MNNIRVSLGPIPPVNPSPW--FLILGLVLTISYGODIARFPQJSGGTP---WILGLIT 64  
QY 56 AVMLIASLAEHLAAVAVARRAGVSVESVTLMLFGCVATLAGEAKTPKAAFRIPAPGAT 115  
DB 65 ALLLPASVVAHELGSLVALAGIEVKSITLFLFGGLASLESSESTPPOAFVAVLAGAV 124  
QY 116 SLASATFGALATITLAVGTPPAIV--ISVAMVLTATVNLGLFNLLPGAPLDGRLVAV 173  
DB 125 SLVL-----FLGLTIVGTQIPLPVPQAIIIGLIGMINIALFNLIPLGLDGGVNLKSI 179  
QY 174 LMRHSDSVRAGIGAGARAVVALVIALGLAEFVAGLNGVWLAIFGMPFPAAREE 233  
DB 180 VMQITGNQNGKILASRVGGGFWLAIIGSIGLITLIPGSEFWTILIGWFLONAGSSA 239  
QY 234 TRISTQCLFAGVRVADMTAQPTAPGWINVEDFIQRYVLGER--SAVPVADRDGSITGL 292  
DB 240 RNAQVKEQMEAFVADAVIENSPIIPAGLNIREFANDVIKTEPRRRLVIGADNQLIGV 299  
QY 293 VALRQLRDVAPSRSTSVGDIALPLHSVPTAR--POEPLTALLERMAPLGRSRALVTEG 351  
DB 300 LATEDIK-----HVPTSDMPQVTVDSLMQY-----PQWVTVVAN 334  
QY 352 SAVVGIVTSPDVARIIDVYRLAQ 374  
DB 335 QSLF-----EVAQLLDQKLSE 351

## RESULT 14

Q7VE60 PROMM PRELIMINARY; PRT; 422 AA.

AC Q7VE60; 01-JAN-1998 (TREMBLrel. 25, Created)  
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Hypothetical protein.  
GN OrderedLocustNames=PW1225;  
OS Prochlorococcus marinus (Strain MIT 9313).  
OC Bacteria; Cyanobacteria; Prochlorales; Prochlorococaceae;  
OC Prochlorococcus.  
OX NCBI\_TaxID=74547;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22825698; Pubmed=12917642; DOI=10.1038/nature01947;  
RA Rocap G., Larimer F.W., Lamerdin J.E., Malfatti S., Chain P.,  
RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,  
RA Johnson Z.I., Land M.L., Lindell D., Post A.F., Regala W., Shah M.,  
RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,  
RA Webb E.A., Zinser E.R., Chisholm S.W.;  
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic  
RT niche differentiation."  
RL Nature 424:1042-1047(2003).  
DR EMBL; BX572098; CAE21400.1; -: Genomic\_DNA.  
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR008915; Peptidase\_M50.  
DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
DR Pfam; PF02163; Peptidase\_M50; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 422 AA; 45664 MW; 7C0E619FDD1F948 CRC64;

Query Match 21.3%; Score 424; DB 2; Length 422;  
Best Local Similarity 29.5%; Pred. 424; 3.1e-21;  
Matches 127; Conservative 69; Mismatches 177; Indels 58; Gaps 10;

QY 1 MRDAIPGIRINGFVNVNHSVLTITLFTMSLATMNLPGATVGVPAVVYLLGAGAVML 59  
DB 1 MGEQWELMKIRIGIPLRVHPSWVFIILLFTWISQNVSAASSLPAMWSGILGTLALL 60  
QY 60 LASLAEHLAAVAVARRAGVSVESVTLMLFGCVATLAGEAKTPKAAFRIPAPGAT 119

DB 61 FLVLAEHLGSLVALAEQVKRSITLFLGCVASVERECSTPMASRVAAGPLVSLVL 120  
QY 120 SATFGALATITAGY---KTPAVISVAMVLTATVNLGLFNLLPGAPLDGRLVAVYLW 175  
DB 121 -----AAVALTGVAADHVNPLANTVGOIGLGNLILALFNLLPGPLDGGILTKALW 175  
QY 176 RRHSDSVRAGIGAGARAVVALVIALGLAEFVAGLNGVWLAIFGMPFPAAREE 233  
DB 176 QWTS-SQKQGVQVATGCRALSSAMVGLGLLFFKGGIGGLMLLMGFGAGASRQT 234  
QY 234 TRISTQCLFAGVRVADMTAQPTAPGWINVEDFIQRYVLGERHSA---YVADRDGSIT 290  
DB 235 QLLALQVLAELNANGQAAGNFRVLDDQSLRLSQRALSGSERQSQEWLVGRSRKW 294  
QY 291 GLVALRQLRDVAPSRSTSVGDIALPLHSVPTARPOEPL----- 330  
DB 295 GYMTDQLKEPLPVQWDRQCLADHMKRISLIPALGERAPQWMALEQAEGRLLVFNV 354  
QY 331 -----TALLEMAPLGR-----SRALVTEGSNAVVGITPBDVARIIDVYR 371  
DB 335 AGLPQGLDRIDSEAVLKR--LGVRLPVQFLPAARQNTYPLGMALPVVESWVSGGL 411  
QY 372 LAQPEPTTS 382  
DB 412 VEQPEASSSTS 422

## RESULT 15

O29915 ARCFU PRELIMINARY; PRT; 362 AA.

AC O29915; 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Hypothetical protein AF0332.  
GN OrderedLocustNames=AF0332;  
OS Archaeoglobus fulgidus.  
OC Archaea; Buryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=96049343; Pubmed=9389475; DOI=10.1038/37052;  
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Ketchum K.A., Dodson R.J., Gwinn M.L., Hickey E.K., Peterson J.D.,  
RA Richardson D.L., Kachavush J., Lee N.H., Sutton G.G., Gill S.R.,  
RA Fleischmann R.D., Quackenbush J., McKenney K., Adams M.D., Loftus B.J.,  
RA Kitzman E.F., Dougherty B.A., McNeill L.K., Badger J.H., Glodek A.,  
RA Peterson S.N., Reich C.I., Weidman J.F., McDonald L.A.,  
RA Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., Karp P.,  
RA Usterback T.R., Cotton M.D., Spriggs T., Artlich P., Kaine B.P.,  
RA Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C.,  
RA Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O.,  
RA Moese C.R., Venter J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-  
RT reducing archaeon Archaeoglobus fulgidus."  
RL Nature 390:364-370(1997).  
DR EMBL; AE001082; AAB90905.1; -: Genomic\_DNA.  
DR PIR; D69291; D69291.  
DR TIGR; AF0332; -;  
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR000644; CBS.  
DR InterPro; IPR008915; Peptidase\_M50.  
DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
DR Pfam; PF00571; CBS; 1.  
DR Pfam; PF02163; Peptidase\_M50; 1.  
DR SMART; SM00116; CBS; 2.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 362 AA; 39865 MW; A09CF565765D7662 CRC64;

Query Match	21.0%;	Score 418;	DB 2;	Length 362;
Best Local Similarity	30.5%;	Pred. No. 6.8e-21;		
Matches 113;	Conservative 70;	Mismatches 162;	Indels 26;	Gaps 10;

Qy	1	MRRAIPGRAGIAGVAVVNMHWSVLVILMLFTWMSLA-TMLPGTGGVPAVVYVYMLLAGGAVML	59
Db	1	MRASFRIKFRVGGIDVEVHSLILILVLLIYASVSPFRYGRANPEPLSRRTITLSSMAAVGL	60
Qy	60	LASLLAHLAHLAAVVARRAGVSVEVYTLMLFGGVTLAGEAKTPKAFAPIAAGPATSLAL	119
Db	61	FASILLAHLEGHSLVARVYGRIRGIMLFIFGVVANMDELPPKPRELVALISGPATS---	117
Qy	120	SAFFGALATTLAGVRRPFAIVSYAAWKL-ATVWLLIGLFNLLPGAPLDGGRLVARYLMRRH	178
Db	118	---FGIAVVSALLSSIPVAELSAFFLLFGYLMFILAIFNLLPAPFMODGRILRSFLAEKR	174
Qy	179	GDSVPAIGIGARAGRVRVALVLTALGLAEPAVGALGVWMLFICGFIIPAAAREETRIST	238
Db	175	SYAEAKTI-NAELGRLLAIFMAIF-----GITTFMILLILFVYIGANNEERVLVL	225
Qy	239	QQLPAGVRVADAMTAQPTAPGWINVEDEFIORYVLGEBHSAPVADRDGSITGLVALROL	298
Db	226	ENVLGRVRADVNMTEVVTTPPEMTYSEVID-LIKTGHLEPFPVEGS-RLVGIITLHDI	283
Qy	299	RDVAPERRSTTSVGDIALPLHSHVTPPARPOEPIFALLERMAPLGGPRRALVTESGAVYGIY	358
Db	284	IGVEPEER-----VGNIM-SREVVAVSPQSAFEAFKIMSEWG-IGRLPVEHGRVVGIV	336
Qy	359	TPSDVARLIDV	369
Db	337	SRSDDLAKIKEI	347

Search completed: March 23, 2006, 05:24:28  
Job time : 102.515 secs



This Page Blank (uspto)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using SW model

Run on: March 23, 2006, 05:24:53 ; Search time 26.8039 Seconds  
(without alignments)  
1212.194 Million cell updates/sec

Title: US-10-617-038-28

Perfect score: 1950  
Sequence: 1 MRDIPILGRIGFVVNVHMS.....QPEPTTSPQDADRFS DAG 393

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/6\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/H\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/RTUS\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/RE\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	244	12.3	290	US-09-902-540-16040	Sequence 16040, A
2	174	8.7	346	US-09-902-540-13409	Sequence 13409, A
3	139.5	7.0	367	US-09-252-991A-29289	Sequence 29289, A
4	125	6.3	404	US-09-712-363-257	Sequence 257, App
5	122	6.1	1033	US-09-252-991A-30611	Sequence 20611, A
6	120.5	6.1	457	US-09-605-703B-2350	Sequence 2350, App
7	120.5	6.1	472	US-09-902-540-15926	Sequence 15926, A
8	120.5	6.1	691	US-09-252-991A-19506	Sequence 19506, A
9	120	6.0	508	US-09-489-039A-7541	Sequence 7541, App
10	120	6.0	574	US-09-902-540-10370	Sequence 10370, A
11	120	6.0	866	US-09-902-540-11981	Sequence 11981, A
12	119	6.0	332	US-09-902-540-15291	Sequence 15291, A
13	116	5.8	810	US-09-902-540-12414	Sequence 12414, A
14	116	5.8	1402	US-09-712-363-166	Sequence 166, App
15	115.5	5.8	551	US-09-198-452A-653	Sequence 653, App
16	115.5	5.8	551	US-09-438-185A-616	Sequence 616, App
17	115.5	5.8	913	US-09-252-991A-29362	Sequence 29362, A
18	115	5.8	312	US-09-252-991A-24594	Sequence 24594, A
19	115	5.8	594	US-09-252-991A-27335	Sequence 27335, A
20	114.5	5.8	592	US-09-902-540-10635	Sequence 10635, A
21	114	5.7	339	US-09-902-540-11039	Sequence 11039, A
22	114	5.7	3724	US-08-804-227C-10	Sequence 10, App1
23	114	5.7	3724	US-08-804-198-4	Sequence 4, App1
24	113.5	5.7	487	US-09-603-208A-238	Sequence 238, App
25	113	5.7	426	US-09-252-991A-25192	Sequence 25192, A
26	113	5.7	661	US-09-134-001C-4303	Sequence 4303, App
27	113.5	5.7	427	US-08-311-731A-216	Sequence 216, App

28	112.5	5.7	562	US-09-902-540-10175	Sequence 10175, A
29	112.5	5.7	625	US-09-438-165A-346	Sequence 346, App
30	112	5.6	411	US-09-252-991A-31301	Sequence 31301, A
31	112	5.6	492	US-09-605-703B-1846	Sequence 1846, App
32	112	5.6	564	US-09-252-991A-17681	Sequence 17681, A
33	112	5.6	3519	US-09-428-517-4	Sequence 4, App1
34	112	5.6	7257	US-09-335-409-5	Sequence 5, App1
35	112	5.6	7257	US-09-568-102-5	Sequence 5, App1
36	112	5.6	7257	US-09-567-969-5	Sequence 5, App1
37	112	5.6	7257	US-09-568-480-5	Sequence 5, App1
38	112	5.6	7257	US-09-568-486-5	Sequence 5, App1
39	112	5.6	7257	US-09-568-472-5	Sequence 5, App1
40	112	5.6	7257	US-09-567-899-5	Sequence 5, App1
41	112	5.6	7257	US-10-014-717-5	Sequence 5, App1
42	111.5	5.6	800	US-09-902-540-11684	Sequence 11684, A
43	111.5	5.6	2910	US-09-828-498-2	Sequence 2, App1
44	111	5.6	953	US-09-902-540-14681	Sequence 14681, A
45	110.5	5.6	534	US-09-252-991A-30725	Sequence 30725, A

## ALIGNMENTS

```
RESULT 1
US-09-902-540-16040
; Sequence 16040, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(11549)B
; CURRENT APPLICATION NUMBER: US/09/902, 540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217, 883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 16040
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
; US-09-902-540-16040

Query Match      12.3%; Score 244; DB 2; Length 290;
Beet Local Similarity 30.5%; Pred. No. 2.6e-16;
Matches 91; Conservative 48; Mismatches 117; Indels 42; Gaps 11;

QY      88  LFGGVTLGGEAKTPKAFRIAPAGPATSLALSTFGALATITLAGVTPAIVISVAMWLA 147
      2  MGVGSSELTAPPPRRDRALMAAVGPELTSLILAVLGATMLDGEVRFNQFAF-FYMA 60
      148  TTNVLLGLENNLPGAPLDGRLVAVYWRHGDVSRAIGAPARAVVALIALGLAEF 207
      61  SLNLFLEGVNLLPAPFMQGRIVRASLAGRLG-WVRATQVSWLGRGPAVLF----- 111
      208  VAGLVGVW-----LAIQWTFPAAABEERFISTOOLFAGVPRADAMTAQPHRAP 259
      112  -----GVMVVLISNPELVIAFFIFMGAGGEAQOVKMTLRRVPADMTTPRRVGD 164
      260  GMINVEDRI-----QRYVIGRHSAYPAVDGSLTGALVALROLDVAPSRSTTSVGD 314
      165  AGASIEQALMLRRAKLL-----LVYTE-DEKPVGVSLIETVRAVDSERMTRTVAV 217
      315  ALPLHSVTPARQPELTALLERMAPLGRSRALVTESAAV-GIVTBSDAVLIDVVR 371
      218  MVP--AVVVRUDEBDGWTA-LRRMAE-BERRPQLVVEADGVLAGTLDVNDVGRMALYQ 271

RESULT 2
US-09-902-540-13409
; Sequence 13409, Application US/09902540
```

Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 13409  
; LENGTH: 346  
; TYPE: PRT  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-13409

Query Match 8.7%; Score 174; DB 2; Length 346;  
Best Local Similarity 27.4%; Pred. No. 4,3e-09;  
Matches 92; Conservative 37; Mismatches 137; Indels 70; Gaps 17;

QY 58 MLASLAEHLAAVAVARRAGVSVTLMFGVTALGGEAK-TPKAAFRAPAPGPTS 116  
DB 19 IVPASVILHEIGHALAFRRYGCPR-IEHGMGGTQTHDAHLTHQSAFVSFAPGIG 77  
QY 117 LALSFGALA-----ITLAVRTPAIVISVAMWLATVNLGLFNLLPGAPADG 166  
DB 78 FLAGLIGLSQVPLGSPGLADGVR-----QFLM-----VINGKLFNLLPQPDG 127  
QY 167 GR---LVRAYLWRHGDVSRAGIGARVALVLIALGAEFVAGLVGWLAFIG 222  
DB 128 GHLADLYVARASGYHERGV-LGVG-----IATAVVLGIAIW-SKQMMGMAMVIG 178  
QY 223 WFTPAABEERTISTQOLFAGVRRADMTAOPHTAP-GWIVNEFIRYVLGERHSAYP 281  
DB 179 VWNLEQLRTPRRRPAERFALPR---LVKPEAPAGALISEQLMDEL---RGTRP 230  
QY 282 --VADRDSITG---LVALRQLRDVAPSRSTSVGDIALPLHSVPTARQEPPL--- 330  
DB 231 PGADADDDLEGPDPALVGENLDNGIPE---LAVSLSQAFQPLATRGHALVLAL 286  
QY 331 --TALLERMAPLGPSSR-----ALVTEGSAVY 355  
DB 287 LHTGRLGELASGLDSSSHARQLSEDTLALISEHAGTV 322

RESULT 3  
US-09-252-991A-29289  
; Sequence 29289, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 29289  
; LENGTH: 367  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29289

Query Match 7.0%; Score 139.5; DB 2; Length 367;  
Best Local Similarity 24.9%; Pred. No. 1.5e-05;  
Matches 98; Conservative 48; Mismatches 143; Indels 105; Gaps 18;

QY 24 LMLFTSLATMLPGTVGYPVAVY-----WLGGAGAVMLASLAHEHLAAVARRAG 78  
DB 27 VTMRLYW--ATPLVAGLGGWLASLVGWPMPMTISLAVMLVTRCLADLPAAEPGARKCG 84  
QY 79 VSVESVTLMLFG-CVTLALGGEAKTPKAAFRAPGPTSLASA--TEGALATLAVRT 135  
DB 85 -----QWIVGIGI-----GLHFPVAVIEQV-----LAHSVIVFGAVATLSSVLA 125  
QY 136 PAIVISVAMWLATVNLGLFNLLPGAPLDGRLVRAVYWRHGDVSRAGIGARRVV 195  
DB 126 IAFMRSGEDRAITFF-----ASMFGASSEMNLGQRHG-AVLSVAAAQSLRL 174  
QY 196 ALVILALGLAEFVAGLVGV-----WLAFF-----IGFTPAABEERT-- 235  
DB 175 LVVLVLPAAFCYLLGGQCPGPHQAAPVDMWMLALLFPAGALVALGW-----QKLRP 226  
QY 236 ----ISTQOLFAGRVA-DAMTAOPHTAPG---WINVEDFIQRVYVGBERRSAYPVARDG 287  
DB 227 NPMLLGPLLLAAGVSLGPDHLHIGLPAGSSGVGWL-----IG 263  
QY 288 SITGLVALROLRDVAPSRSTSVGDIALPLHSVPTARQEPPLT-----ALLERMAPLGP 342  
DB 264 SALGCHNRSPFRAPAFVERTIVCTLMMFPAALAAELGWLTTLDHQSIMLGMMCGI 323  
QY 343 RSRALVTEGSAV-VGIPTSDVARLIDVYRLAQP 375  
DB 324 AELSLTAEALQLSVPLVTAQLVRLLLVLFLAEP 357

RESULT 4  
US-09-712-363-257  
; Sequence 257, Application US/09712363  
; Patent No. 6892139  
; GENERAL INFORMATION:  
; APPLICANT: Eisenberg, David  
; APPLICANT: Rotschein, Sergio H.  
; APPLICANT: Marcotte, Edward M.  
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
; FILE REFERENCE: 07419-032001  
; CURRENT APPLICATION NUMBER: US/09/712,363  
; PRIOR FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: PCT/US00/02246  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: 60/179,531  
; PRIOR FILING DATE: 2000-02-01  
; PRIOR APPLICATION NUMBER: 60/117,844  
; PRIOR FILING DATE: 1999-01-29  
; PRIOR APPLICATION NUMBER: 60/118,206  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: 60/126,593  
; PRIOR FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: 60/134,093  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/134,092  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/165,124  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/165,086  
; PRIOR FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 257  
; LENGTH: 404  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-712-363-257

Query Match 6.3%; Score 125; DB 2; Length 404;  
Best Local Similarity 28.9%; Pred. No. 0.00051;  
Matches 57; Conservative 20; Mismatches 66; Indels 54; Gaps 9;



```
; NUMBER OF SEQ ID NOS: 16625
; SEQ ID NO 15926
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-15926

Query Match      6.1%; Score 120.5; DB 2; Length 472;
Best Local Similarity 22.9%; Pred. No. 0.0019;
Matches 92; Conservative 53; Mismatches 133; Indels 123; Gaps 21;

QY 38 GTVGVPVAVVWMLGAGGAVMLIASL--AAHELA-----HAAVBARAGVSVESTL 86
DB 107 GSMIGHPAVV--TGLAVFLILLATSMFGAFELAPSSFOTKLNAVGG--AGVA---GA 157
QY 87 WLFPGVTA-LGGEAKTPRAAFRIAAGAPATSLASATFGALATITLGAIVTAIVSV--- 142
DB 158 FLMSVSGFLAPCPGFLVTGLLNAVAKSANTTLGASL--LFYVALGIGVFPLIGVTV 215
QY 143 -----AMWLATVNLILGLFNLIPGAPLDGRLVRAVYLRHRGDSVRA-GIGAAARA-GRV 195
DB 216 RLPRGVMMWVWKSVTGIMLV---ALAFSYLKDAFPWAR--DVYKGLGLHVGRVPAGVI 269
QY 196 ALVLIAGLA-----ER-----VAGGLVGVWLAFIG 222
DB 270 AALLVAVGVVALGAVHRSEFKEGAREFSFALGVALLVALLRGALDGGVGTLMVS-LG 328
QY 223 WFIPLAAEEETRISTOOLFAGVAVADMTAQPHTAPGM-----INVEDFIQR 270
DB 329 -----LAEPARAPMCHHWAPAKQATPSPEAPDQ- 358
QY 271 VYIGERSAVVADRDGSIITGLVALRQL-RVAPERSSTSVGDIALPLHSVPTARPOEP 329
DB 359 -VLAQAKKEGRPVLIIDFPADWCACKELDRDTPYPAQVYSSEDEGRFINIKIDATNSDS 417
QY 330 LTLAERMAPLGRPSRALVT-EGSA-----VVGIVTPSDVA 364
DB 418 LDALLERFVEGLPTVAFTSPEGKVLTPQPRVTFGLAFSPFA 458

RESULT 8
US-09-252-991A-19506
; Sequence 19506, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107136.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19506
; LENGTH: 691
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19506

Query Match      6.1%; Score 120.5; DB 2; Length 691;
Best Local Similarity 25.7%; Pred. No. 0.0033;
Matches 106; Conservative 40; Mismatches 166; Indels 101; Gaps 20;

QY 50 LLAGGAVMLIASLAEHLAAVAVARRA--GVSVESVTLMFGVTTALGGEAKTP----- 102
DB 78 VLAHDHALGVVALGTGELGALRLRVPGIGQLGAVAGLHGAN--VGCTQPAFLRVLV 135
QY 103 -KAAPRIAFAGPATSLASATFGALATITLAVR-----TPAIVISVA---MWLAVTNLL 153
DB 136 GRVEDRVGIALPVVA-AVPARGGQOATTFLLAVRIASGDLFPADVLFVGEHDVQRHANLAG 194
```

```
QY 154 GLFNLPFGAPLDGRLVRAVYLRHRGDSVRAGIG---AARAGVVALVLIAGLAEFVA 209
DB 195 GVEGLIDGL---GPRQVEAVVOGLVGFLLDAPVQSVALLQRORLAEVDLALAOGAVQORA 251
QY 210 -GGLVGV-----WLAFIGFIFAAREEETRISTOOLFAGVAVADMTAQPHTAP 256
DB 252 MAGFVVDLAPRLALDEEGFLAFAGWLPLAVGVQIEVALASQLF---QVEILLIVQDHC 308
QY 257 TAPMINVEDFIQRVYIGERSAVPVADRDGSIITGLVALRQLRDVARSSTTS---VG 312
DB 309 HAPQLAVE-----ABHHARQTS--DGNARGVLGRADLHVVPBRRHAQRQGVVYG 357
QY 313 DIAL-----PLHSVPTARPOEPTLTLERMAPLGRPSAL--VREGSA 353
DB 358 HQALAAALRRDSPVVRGGDPBHAQRCDLPRRALDREAGNLPVQAQALEPRLVERQV 417
QY 354 VVGI-----VTPSDVAR-----LIDVYRLAQP-----PTFTT 381
DB 418 FVGIQRQPGQLVAADLLERQRQDPFLQVHRQAEEVQEGEDQGRVLDLPVFRT 470

RESULT 9
US-09-489-039A-7541
; Sequence 7541, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7541
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7541

Query Match      6.0%; Score 120; DB 2; Length 508;
Best Local Similarity 23.6%; Pred. No. 0.0023;
Matches 86; Conservative 47; Mismatches 140; Indels 92; Gaps 18;

QY 10 IAGFVVVWVH--SVYVILMLFTWSLATMLP-GTVGGVP----- 44
DB 183 VGGFLA-VSMTDVTYQASLMIFALITFTVYIIVSGGSDSLEVIKKSINIDMLKGLNF 241
QY 45 -AAVYVWLLGAGGAVMLIASLAEHLAAVAV-ARRAGVSVESTLMT--FGVTTALG 96
DB 242 VAISLWGMGLGFQGCQHIIILAFPAADSHSHSIYHARRISMT-----WMLICGGAAYVG 295
QY 97 GEATTPRAAFRIAAGATSLASATFGALAI-----TLAGVTPAIVISVAVW 145
DB 296 -----FFGIAFFNNPSLAGAVNONARVFTELAQIILFNPWIAIILISALIAV--- 344
QY 146 LATVNLILGLFNLIPGAPLDGRLVRAVYLRHRGDSVRAGIGAAARAGVVALVLIAGL- 204
DB 345 MSTISC-----QLIVCSATITEDLYKAFKRNQOKELVWVGKRMV-LVALVALMALAAN 398
QY 205 AEFVAGLVGVWLAFIGWF---IFAAREEETRISTOOLFAGVAVADMTAQPHTAPG 260
DB 399 PENRVGLAVSVAMVFGAAGFPVVLFFVMSRMTR---NGALAGM-VIGALTIVWKRQFG 454
QY 261 WINVEDFIQRVYIGERSAVPVADRDGSIITGLVALRQLRDVARSSTTSVGDIALPLHS 320
DB 455 WLGXEYIIPGVFSG-----IGIVVF-SLIDKAPASWQORFAEDAHYHT 499
QY 321 VPTAR 325
DB 500 PEPVR 504
```

RESULT 10  
US-09-902-540-10370  
; Sequence 10370, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR FILING DATE: 2000-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 10370  
; LENGTH: 574  
; TYPE: PRT  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-10370

Query Match 6.0%; Score 120; DB 2; Length 574;  
Best Local Similarity 22.2%; Pred. No. 0.0028;  
Matches 95; Conservative 53; Mismatches 126; Indels 154; Gaps 23;

QY 11 AGFVNVNH--WSVLV--ILMLFTSLATMLPCTVG--GYPAVVYLLGAGCAVMLLAALL 64  
DB 204 AGALVSFPTGWSMLGAVLTGFLAPAVWSQALGEVTRAINSMVMTGSAVLVSSGLL 263  
QY 65 AHELAHAVABRAGVSVETLMLFGVTALGGEA-----KTPKAERIAFAGPATSLA 118  
DB 264 SFAQMRVAR-----SFKSLG--GLFKGKAKEDMDPLAGIECPRAWPLG----- 310  
QY 119 LSATFGALATITLAVRTPAIVISVAMWLATVNLGLFNLPGADLDGRLVRAVLMRRH 178  
DB 311 --AIGPVAVFML-----AYLFQIPMWAGVLM-----PL----- 338  
QY 179 GDSVRAGGAAR-AGRVVALVILAG--LAEPVAGCL-----VGGVWLAFI- 221  
DB 339 --AVMGVIASRVGTETDTPTKALGPVTQLFGSLAPGNI PANVMSANATGGVGLHSAD 396  
QY 222 -----GMFPAAREETRISTQQLF--AGVAV-----ADAMTAOPHTAPG 260  
DB 397 LITDLKSGMLGANRQ-----FVAQLFGVAGAAVVPVKIIVPDASMLGTEFPAPA 452  
QY 261 --WINVEDFIOQVVLGERHSAVPAVDRDGSITGLVALRQLRDVAPSRRTSTVGDIAP 317  
DB 453 SMVAVGVSKLATGV-----AALPESARWGLCG-----AALG 485  
QY 318 LHSVPTARPOEPLTLLERMAPLGRS-----RALVTEGSAVVGIVTPSDVARIID 368  
DB 486 IFLV-----LLERMAPAKAKAVVSPBAGGLAIVIPGSSSIAFPGLSAIAELL- 533  
QY 369 VYRLAQPE 376  
DB 534 --RRTKPK 539

RESULT 11  
US-09-902-540-11981  
; Sequence 11981, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540

; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 11981  
; LENGTH: 886  
; TYPE: PRT  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-11981

Query Match 6.0%; Score 120; DB 2; Length 886;  
Best Local Similarity 23.0%; Pred. No. 0.0054;  
Matches 101; Conservative 44; Mismatches 149; Indels 145; Gaps 24;

QY 46 VVWMLGAGAVMLASL-----LAHELAHVA-----RRAGVSVESVT 85  
DB 235 VAVVILASGAVVAGALGAFSSRPGDPQVDLASALAPILAAGVFLPEBRRLVHLMMT 294  
QY 86 L-----WIFGVTALG-----EAKTPKAERI--AFAPATSLA 118  
DB 295 TILGALFARTQAPTPDQMWVFGGTAVTAGIMPLARWSAPFGLRIRLAWTILGSLVAMP 354  
QY 119 LSATFGALATITLAVRTPAIVISA-----W-----WLATVNLGLFNLPGAP-LDG 166  
DB 355 LVSOVGSASFPQVLTGLVITAGAAHVAGKRWKGLHYGLFGVLAFFVEGTPALEG 414  
QY 167 GRLVRAVLMRRHGDVBRAGIGABARGVVALVILALGLAEFVAGLVGCVWLAFIGWFI 226  
DB 415 S-----W-----GRLAVPALVS--GL--YGVAGLVGAMA----- 440  
QY 227 AAAREETRISTQQL--FAGRVADAMTAOPHTAPGMINVEDFIQVILGERHSAVPA 283  
DB 441 OPASRBDGLPLDIDLALCLAAAGVTLAVDSSP--SAPDLVSVTGLSASVL-----SALPFA 495  
QY 284 -----DRDS--ITGLVA-----LRQLRDVAPSRRTSTVGDIAPLHSV 321  
DB 496 VVTAMLLRVRDSSRLVGFLLAASGLAAVSQVGTSDFA--SPRAALVAASLALGEFAF 554  
QY 322 PTAPOEPLTAL-----LERM-----APLGRSRLVTEGSAVVGIVTPSDVARIIDV 369  
DB 555 SALAGRTPDLADADTAAGERRRMDLRLPLGARGPLFTDGFATLALV-QTVIALTLA 613  
QY 370 YRLAQPEPTTSPQADR 388  
DB 614 GMLAR-----PTDAER 624

RESULT 12  
US-09-902-540-15291  
; Sequence 15291, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR FILING DATE: 2000-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 15291  
; LENGTH: 332  
; TYPE: PRT  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-15291

Query Match 6.0%; Score 119; DB 2; Length 332;  
Best Local Similarity 26.4%; Pred. No. 0.0016;  
Matches 64; Conservative 19; Mismatches 71; Indels 88; Gaps 13;



```
OY 281 PVADRDGSTITGLVALKQLRD-----VAPSRSTTSVGDIALPL-----HSVPT----- 323
DB 1030 -VLGRDGA-AGLDAGQAFADSGFNSLSAVELNRRLTAATVATLPAFAIFDHPTELAQY 1087
OY 324 -----ARPEPLTALLE-----RMAPLGRSR----- 345
DB 1088 LITQIDHGSSAAAAANPARIIDALTDLFLOACDAGRADGKMWALASNTRERMSSPVR 1147
OY 346 -----ALVTEG-SAVGVITPS-----DVARLI-----DVRRLAQBPPTTS 382
DB 1148 NNVSXKVALADGIDVDVITCIPTLTVLSDQREYRDIANAMTGRHSVSLTL--REFDSS 1205
OY 383 ---PODAD 387
DB 1206 DALPONAD 1213

RESULT 15
US-09-198-452A-653
; Sequence 653, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Giffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 653
; LENGTH: 551
; TYPE: PRF
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-653

Query Match 5.8%; Score 115.5; DB 2; Length 551;
Best Local Similarity 21.4%; Pred. No. 0.0075;
Matches 65; Conservative 40; Mismatches 94; Indels 105; Gaps 14;

OY 1 MRDAIPLRIGAFVNVNH-----MSVLVILMLFTWSLALMLP--GTVGGYPA 45
DB 140 LQELLPQG-LRGFLVMVRWYSYIYVMSLSVVLSTMLF-WGLANQITITTEAGRFPYA 197
OY 46 VV-----VYLLGAGAVMLASLAEHLAAVAVRAGVSVE--VTLM 88
DB 198 LINTLNLSSICAGEISTWM---GKQTFVAISFACDSHWSVNLMLTITCSGLIMIML 253
OY 89 F-----GGVTALGCEAKTPKAAFRIFAAGPATSLALSTPGAL 126
DB 254 YRRIHHLITDISIPPSRRLTAEGEAATNLKKEKKPKAKARLNF---LHLQSRVYLGL 309
OY 127 AITLAGVTRPAIVISVAV-----WLAIVNLLGLFNLPGAPLDGRLV 170
DB 310 AITVSYNLVHLFEVVMKQVQSQIYSHVEFNWMSRITTLIGVSVL-AAVLLTGQCI 368
OY 171 RAYLRRHODSVRAGIGARARVVALVLIAGLAEF-----VAGGLVGWMLA 219
DB 369 RKMGMT-----VGLA---VTPVLVSGLLFETIFAKRDISIFGVLGMTPLA 415
OY 220 FIGW 223
DB 416 LAAW 419
```

Search completed: March 23, 2006, 05:30:18  
Job time : 28.9039 secs



This Page Blank (uspto)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 06:14:33 ; Search time 82.6454 Seconds  
(without alignments)  
1986.885 Million cell updates/sec

Title: US-10-617-038-28

Perfect score: 1990  
Sequence: 1 MRDAPLGRAGFVNNVHWS.....QPEPTTTSPODADRFSFDAG 393

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA Main:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1990	100.0	393	4	US-10-617-038-28 Sequence 28, Appl
2	425.5	21.4	379	4	US-10-047-260-12 Sequence 12, Appl
3	311	15.6	376	4	US-10-156-761-14201 Sequence 14201, A
4	143	7.2	473	4	US-10-156-761-14437 Sequence 14437, A
5	140	7.0	591	5	US-10-482-706-183 Sequence 183, App
6	140	7.0	649	4	US-10-080-170-24 Sequence 24, Appl
7	140	7.0	649	4	US-10-080-170-24 Sequence 24, Appl
8	140	7.0	649	4	US-10-468-356-24 Sequence 24, Appl
9	138.5	7.0	706	4	US-10-282-122A-50265 Sequence 50265, A
10	133.5	6.7	570	4	US-10-156-761-9968 Sequence 9968, Ap
11	133	6.7	944	5	US-10-493-462-25 Sequence 25, Appl
12	132	6.6	3745	4	US-10-205-032-14 Sequence 14, Appl
13	130.5	6.6	3362	4	US-10-378-083-6 Sequence 6, Appl
14	130.5	6.6	4685	4	US-10-156-761-10433 Sequence 10433, A
15	128	6.4	487	4	US-10-437-963-17542 Sequence 17542, A
16	127.5	6.4	400	4	US-10-282-122A-48123 Sequence 48123, A
17	126.5	6.4	851	3	US-09-991-936-1915 Sequence 1915, Ap
18	126.5	6.4	851	5	US-10-978-245-1915 Sequence 1915, Ap
19	126.5	6.4	3808	4	US-10-378-083-7 Sequence 7, Appl
20	125.5	6.3	466	4	US-10-156-761-8852 Sequence 8852, Ap
21	125	6.3	404	3	US-09-712-363-257 Sequence 257, App
22	125	6.3	533	4	US-10-424-599-215899 Sequence 215899, A
23	122	6.1	1029	4	US-10-389-647-613 Sequence 613, App
24	121.5	6.1	477	4	US-10-425-115-310459 Sequence 310459, A
25	121	6.1	833	4	US-10-156-761-12366 Sequence 12366, A
26	120.5	6.1	450	4	US-10-156-761-11177 Sequence 11177, A
27	120.5	6.1	457	3	US-09-738-626-4139 Sequence 4139, Ap

28	119	6.0	413	4	US-10-282-122A-62054 Sequence 62054, A
29	119	6.0	811	5	US-10-450-763-59793 Sequence 59793, A
30	118	5.9	791	4	US-10-156-761-12134 Sequence 12134, A
31	117.5	5.9	323	4	US-10-156-761-8143 Sequence 8143, Ap
32	116.5	5.9	260	4	US-10-282-122A-61660 Sequence 61660, A
33	116.5	5.9	531	4	US-10-437-963-158999 Sequence 158999, A
34	116	5.8	1402	3	US-09-712-363-166 Sequence 166, App
35	116	5.8	7257	5	US-10-732-923-20621 Sequence 20621, A
36	115.5	5.8	384	4	US-10-156-761-9976 Sequence 9976, Ap
37	115.5	5.8	385	4	US-10-437-963-167452 Sequence 167452, A
38	115.5	5.8	468	4	US-10-425-114-57349 Sequence 57349, A
39	115.5	5.8	540	4	US-10-282-122A-54741 Sequence 54741, A
40	115.5	5.8	551	4	US-10-289-762-653 Sequence 653, App
41	115.5	5.8	561	4	US-10-425-115-516538 Sequence 516538, A
42	115.5	5.8	1061	4	US-10-282-122A-49840 Sequence 49840, A
43	115	5.8	403	3	US-09-976-059-23 Sequence 23, Appl
44	115	5.8	548	4	US-10-282-122A-62037 Sequence 62037, A
45	115	5.8	596	5	US-10-762-107-82 Sequence 82, Appl

ALIGNMENTS

RESULT 1									
US-10-617-038-28									
; Sequence 28, Application USV10617038									
; Publication No. US20040057963A1									
GENERAL INFORMATION:									
; APPLICANT: Andersen, Peter									
; APPLICANT: Rosenkrands, Ida									
; TITLE OF INVENTION: Therapeutic TB Vaccine									
; FILE REFERENCE: SSI5AUSA									
; CURRENT APPLICATION NUMBER: US/10/617,038									
; PRIOR FILING DATE: 2003-07-11									
; PRIOR APPLICATION NUMBER: DK PA 2002 01098									
; PRIOR FILING DATE: 2002-07-13									
; PRIOR APPLICATION NUMBER: US 60/401,725									
; PRIOR FILING DATE: 2002-08-07									
; NUMBER OF SEQ ID NOS: 187									
; SOFTWARE: Patentin version 3.2									
; SEQ ID NO 28									
; LENGTH: 393									
; TYPE: PRT									
; ORGANISM: Mycobacterium tuberculosis									
US-10-617-038-28									
Query Match 100.0%; Score 1990; DB 4; Length 393;									
Best Local Similarity 100.0%; Pred. No. 2.1e-173;									
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MRDAPLGRAGFVNNVHWSVTLVTLMTFTWSLAMLPGTVGGYPAVVYVWLLGAGAWLL	60						
DB	1	MRDAPLGRAGFVNNVHWSVTLVTLMTFTWSLAMLPGTVGGYPAVVYVWLLGAGAWLL	60						
QY	61	ASLAAHELAHVARRAGVSVESTLMLFGSVTLAGEAKTPKAAFRIFAAPATSLALS	120						
DB	61	ASLAAHELAHVARRAGVSVESTLMLFGSVTLAGEAKTPKAAFRIFAAPATSLALS	120						
QY	121	ATFGALAITTAGVTPPAIVISVAMWLTATVLLGLFNLLPGAPLDGRLVPAIYMRHGD	180						
DB	121	ATFGALAITTAGVTPPAIVISVAMWLTATVLLGLFNLLPGAPLDGRLVPAIYMRHGD	180						
QY	181	SVRGIGAAARGRVAVLTALGLAEFVAGLVGWLAFIIGWTFPAAREEETRISTQQ	240						
DB	181	SVRGIGAAARGRVAVLTALGLAEFVAGLVGWLAFIIGWTFPAAREEETRISTQQ	240						
QY	241	LFAGVRVADAMTAQPHTPAGVINVEDFIQRYVLGERHSAYPVADRDGSLTGLVALROLRD	300						
DB	241	LFAGVRVADAMTAQPHTPAGVINVEDFIQRYVLGERHSAYPVADRDGSLTGLVALROLRD	300						
QY	301	VAPRRSTTSVGDIALPLHSVTPARPOEPTTALIERNAPLGPNSRALVTBGSAAVVGIVTP	360						
DB	301	VAPRRSTTSVGDIALPLHSVTPARPOEPTTALIERNAPLGPNSRALVTBGSAAVVGIVTP	360						

Db 301 VAPSRRTSVGDIALPLHSVPTARPQEPETALLERMAPLGRSRALVTEGSAAVGIYTP 360  
Qy 361 SDVARLIDVYRLAOPETFTTSPQDADRFSDAG 393  
Db 361 SDVARLIDVYRLAOPETFTTSPQDADRFSDAG 393

## RESULT 2

US-10-047-260-12  
; Sequence 12, Application US/10047260  
; Publication No. US20020164706A1

GENERAL INFORMATION:

APPLICANT: Huang, Lisa

APPLICANT: McCluskey, Michael

APPLICANT: Labosaa, Robert

TITLE OF INVENTION: HIGH LEVEL PROMOTERS FROM CYANOBACTERIA

FILE REFERENCE: C11715 US NA

CURRENT APPLICATION NUMBER: US/10/047,260

CURRENT FILING DATE: 2002-01-15

PRIOR APPLICATION NUMBER: 60/264,925

PRIOR FILING DATE: 2001-01-30

NUMBER OF SEQ ID NOS: 40

SOFTWARE: Microsoft Office 97

SEQ ID NO 12

LENGTH: 379

TYPE: PRT

ORGANISM: *Synechocystis* sp. strain PCC6803

US-10-047-260-12

Query Match 21.4%; Score 425.5; DB 4; Length 379;  
Best Local Similarity 30.0%; Pred. No. 3.9e-30;

Matches 115; Conservative 70; Mismatches 147; Indels 51; Gaps 11;

Qy 1 MRDAIPLRIG--FVVNVHMSVLVILMLFTWSLATML---PGTVGGYPAVVYVWLLGAGG 55  
Db 11 MNNNRVSGSLRGIPFYVPSW--FIIIGVITISYQDLARFPQLSGTR---WILGLIT 64  
Qy 56 AVMLASLHAEHLAAVAVARRAGSVESVTLMFGVTTALGGEAKTPKAAFRIFAPAGAT 115  
Db 65 ALLPLASVVAHEHLSVALAGIEVKSITLPLFGSLASLEKESWTQAFVAIAGDAV 124  
Qy 116 SLASATGALATITAGRTPAIV--ISVAMTLAVNLLGLFNLLPGAPLDGGRLVAY 173  
Db 125 SLV-----FLGLITVGTQIPLPVGQALIGLGINIALAFNLIPLGLDGGVLAISI 179  
Qy 174 LMRHGDVSRAGIGARAGRVVALVILGLAEFVAGVGGVWLAFTGMFIFAAREEB 233  
Db 180 VWQITGNGKGLISRVGGGWLALAGSLGINIIPIGSFWTIILGWFLQWAGSSA 239  
Qy 234 TRISTQDPAFGRVADAMTAPHTAPGMINVEDFIQRYVLGER--HSAYPVADRDGSTITGL 292  
Db 240 RNAQVKEQMEAFTAEDAVIPNSPIIPAGINIREFANDVYIGKTPWRREFVIGADNQLGIV 299  
Qy 293 VALQLRDAVSRRTTSVGDIALPLHSVPTAR--QEPETALLEMAPLGRSRALVTEG 351  
Db 300 LATEDIK-----HVPISDWPQVTVDSIMQY-----PQNVTVAN 334  
Qy 352 SAVVGIVTPSDVARLIDVYRLAQ 374  
Db 335 QSLF-----EVAQLDDQKLSE 351

## RESULT 3

US-10-156-761-14201

; Sequence 14201, Application US/10156761

; Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 14201  
; LENGTH: 376  
; TYPE: PRT  
; ORGANISM: *Streptomyces avermitilis*  
US-10-156-761-14201

Query Match 15.6%; Score 311; DB 4; Length 376;  
Best Local Similarity 31.0%; Pred. No. 1.2e-19;  
Matches 124; Conservative 47; Mismatches 173; Indels 56; Gaps 16;

Qy 7 LGRIAGFVVNV--HMSVL--VILMLFTWSLATMLPGTVGGYPAVVYVWLLGAGAVMELLAS 62  
Db 1 MGRPFQVTVVYVAPSMFLVAALITWVFGQLDRVLPEL-----GAARYLVSLFFAVAFVAS 55  
Qy 63 LIAEHLAAVAVARRAGSVESVTLMFGVTTALGGEAKTPKAAFRIFAPAGATSLASAT 122  
Db 56 VLIHELHATVAALPFKLPVRRIQLQFQGVSEIKETETPGRFEVLAFFVGLSLIAGI 115  
Qy 123 FGALATITAGVTRTAPVIVSVAMWTLATVNLGLFNLLPGAPLDGGRIVRVRMRHDSV 182  
Db 116 FYLMQVIVDKTGVGVLLA---GLMISNLVAAFNLLPGLPLDGGRLRAVVMKITGKPM 172  
Qy 183 RAGIGAARAGRVVAL-VLIALGL-----AEFVAG-----GLVGVWLAFTGM-- 223  
Db 173 SGTIAAAMVGRALAVAVILGLPLTGTGALGSNAEDVGMDVTDALAAI-LAAIITWG 231  
Qy 224 ----FTFAAREESTRISTQQLFAGVAVADAMTAPHTAPGMINVEDFIQRYVIGERHSA 279  
Db 232 AGNSLRMARLREHLPELRARLTR--RAVPEYDTPLS-----EALRRANDAGAR--A 280  
Qy 280 YFVADRDGSTITGLVALQLRDAVSRRTTSVGDIALPLHS---VPTARQEPETALLER 336  
Db 281 LVVADADGEPLSLREAAIVGPEHRRPWDVSLADLTDGKMSLAGEDLIDVURA 340  
Qy 337 MAPLGRSRALVTEGSA--VVGIVTPSDVARLIDVYRLAQ 375  
Db 341 T-----PATEYLVVESEGEIVGLSAADVERAF-VKAAARP 375

## RESULT 4

US-10-156-761-14437

; Sequence 14437, Application US/10156761

; Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 14437

LENGTH: 473

TYPE: PRT

ORGANISM: Streptomyces avermectilis  
US-10-156-761-14437

Query Match 7.2%; Score 143; DB 4; Length 473;  
Best Local Similarity 27.7%; Pred. No. 0.0004;  
Matches 98; Conservative 36; Mismatches 164; Indels 56; Gaps 17;

QY 67 ELAAHVAVRAGVSVE---VTLMFGVTLAGEAKTPKAFRIAPGATSLASATF 123  
DB 37 EKAAEGRRAGTVESELSFQISG--TQSG---ITTSLVGMLAPALADLHGPF 91  
QY 124 GALATTAGTTPALVISAAMWLAT-VNLLGLFNLLP-----GAPLDGRLVRA--YIM 175  
DB 92 TAIGPEGAV--PGIAVVVGMLASAVQVIG--ELVPKMAISKPELQVAFVAPQHYF 147  
QY 176 RRGDSVRAGIGAAAGRVAVLVLALGL--AEFVAGLVGGVWLAFTGWFIFAAREE 233  
DB 148 SRLFRPVTAALNTV-ANRLVR---ALGVPEEEMASARTPSELVSLARHSQAQLEOD 202  
QY 234 TRISTOOLPAGVRVADAMTAOPHTAP-----GWINVEDEFIOHYVLGERHSAPPV-AD 284  
DB 203 ----TADLFVRTLSIGELTAGHVMTPRVRVSALQSSATAEDEV-NUTRATGLSRFPVYRD 257  
QY 285 RDGSLTGLVALRQLDVAAPSRSTTSVGDIALPLHSVTPARQEPULTALLERMAELGPRS 344  
DB 258 RIDELVGVWHLKDALAIPSRLRTPPVGRIAQPPULVETLFPVQPLARLSEOPIA--- 314  
QY 345 RALVTEGSAVVVGIVTPSDVARLI-----DVRRLAQEPFTTSPDAD 387  
DB 315 -VVDEYGGTAGVTLLEDIVELVGVSRDEHDGQDLPELAVAPEDGRPAMDAD 367

#### RESULT 5

US-10-482-706-183  
; Sequence 183, Application US/10482706  
; Publication No. US20040241826A1  
; GENERAL INFORMATION:  
; APPLICANT: James, Brian William  
; APPLICANT: Marsh, Philip  
; APPLICANT: Hampshire, Tobias  
; TITLE OF INVENTION: Mycobacterial Antigens Expressed During Latency  
; FILE REFERENCE: 1581.1030000  
; CURRENT APPLICATION NUMBER: US/10/482,706  
; PRIOR FILING DATE: 2004-01-02  
; PRIOR APPLICATION NUMBER: PCT/GB02/03052  
; PRIOR FILING DATE: 2002-07-04  
; PRIOR APPLICATION NUMBER: GB 0116385.6  
; PRIOR FILING DATE: 2001-07-04  
; PRIOR APPLICATION NUMBER: GB 0123993.8  
; PRIOR FILING DATE: 2001-10-05  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 183  
; LENGTH: 591  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-482-706-183

Query Match 7.0%; Score 140; DB 5; Length 591;  
Best Local Similarity 24.0%; Pred. No. 0.001;  
Matches 111; Conservative 49; Mismatches 169; Indels 134; Gaps 23;

QY 10 IAGFVVNVH---WS---VLVILMLF-----TMSLATMLDGTVG---GVP 44  
DB 70 VAGAIIRAPLVWTPAPISLVVQLLESIALRALYVSSMRPVLLITPPLPLPLAVP 129  
QY 45 AVVVYLLGAGAVMLASLAHELAAVAVRAGVSVEVTLMFGVTLAGEAKTPKA 104  
DB 130 GFAMWAALNSLPLMAA--LAWVCADAILVLTGHHRAVLT---GVLVYLGGLFFPEKA 183  
QY 105 AF--RIATAGPATSLALSTFGALA-ITLACVR--TPRALVISVAN---MLATV----- 149  
DB 184 AVTPVSPAVVALQCHVGRGDSALATVWRAGVRLMTPSLALTGVAVLAVLVADQRRWS 243

QY 150 -----NLLGLFNLLPGAPLDGRLVRAVLMRRHDSVRAGIGAAPAGRVVA-- 196  
DB 244 DLSMTWDLCSVTHIGVIALAGGFWMDWARWAPSPW-----ATPAAVVWVL 290  
QY 197 --LVIIAGLAEFVAGLVGVWLAFTGWFIFAAREERTISTOQLFAGRVADAMTAQ 254  
DB 291 GMLVILIALALSLVRKRRIIGFWMLTAAG--YAAACQPIFLMSSPFTALELQTLRYF 347  
QY 255 PHTAGWINVEDFIORVYVGERHSAYPVADRDGSLTGLVALRQLDVAAPSRSTTSVGD 314  
DB 348 P-----DLV--VVL-----ALLAAVLAQ-----APNAGTRMLD-- 374  
QY 315 ALPLHSVPT-ARPEPTALLERMAPLGRSRALVTEG---SAVVGIVTPSDVARLID-- 368  
DB 375 ASPAAVATVASAVLFTLSLYSTATFLASWRDNPTEGYLKNQAQSLAAASGAPILDOE 434  
QY 369 -----VRLAQPE-----PFTTSPDADRFSDAG 393  
DB 435 VDPVLQRAVAMPENLASHMFLALRLVRPFAVTTQLRMFTSTG 477

#### RESULT 6

US-10-080-170-24  
; Sequence 24, Application US/10080170  
; Publication No. US20030123601A1  
; GENERAL INFORMATION:  
; APPLICANT: COLE, S. T.  
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR  
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR  
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES  
; FILE REFERENCE: 03495.0218  
; CURRENT APPLICATION NUMBER: US/10/080,170  
; PRIOR FILING DATE: 2002-06-10  
; PRIOR APPLICATION NUMBER: 60/270,123  
; PRIOR FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 652  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 649  
; TYPE: PRT  
; ORGANISM: Mycobacterium leprae  
US-10-080-170-24

Query Match 7.0%; Score 140; DB 4; Length 649;  
Best Local Similarity 23.5%; Pred. No. 0.0011;  
Matches 114; Conservative 62; Mismatches 172; Indels 138; Gaps 25;

QY 11 AGFVVNVHMSVYLILM---LFTWSL---ATMLP-----GTVG 41  
DB 105 AALVLSVLGVWVLMLGTGRLVAPSLQGRQAIMLPAGALVVALPPARDFATSGLESGLVL 164  
QY 42 GYPAVVTYL-----LGAGAVMLASLAHELAAVAVRAGVSVEV 84  
DB 165 TYLGILMWMVMVCMAPLNRQSRRFIGALAFVAGCSVLPBELMLGSGSLIMLMTAAR 224  
QY 85 TLMFGVTLAGEAKTPKAFRIAPAG---PATSLASAT---FGALATITLACVTRPAI 138  
DB 225 TCMRLALIVAGGSLPAVAYQLFRWGYGLVPGTALAKDAGDKMSGCIITLSHPNDPV 284  
QY 139 VISVAMWLATVNL-LGLFNLL-----PGAPLDGR-----L 169  
DB 285 L-----WVPLVLVLVLGLLMLHRMPSFMHPLFTPDGGRVAVAVQSPAVVFFVFGSL 339  
QY 170 VRATLMRR-----HGDSVRAGIGAAAGRVAVLVLALGLAEF--VAGSLGVG---W 217  
DB 340 LQAFYMWIRGGDFMHRGVTLPFLCLAPVVVIVVISSEG--ADFSRQTGNWLAGVTSILW 398  
QY 218 LAFITGFIFAA-----AREERTISTOQLFAGRVADAMTAOP--HTAPGWINVEDFIORY 271  
DB 399 LGVAGKSLMAANSQMGKDDATNVS-----YSGIVDERFRYAQATGHAP--LTAADYL--- 449  
QY 272 VLGERHSAYPVADRDGSLTGLVALRQLDVAAPSRSTTSVGDIALPLHSVTPARQEPBLT 331

```
Db 450 -----GYP---RMAAV--LVALNTPDGLALLPSGNYIKMDLVPMTQLSPSSGSPD 497
Qy 332 ALBERMAPLGRSRALVTESGAVVIGTVPDVARLIDVYRLAQPEPTTSPQ-----D 385
Db 498 SLVSQK----PQHTVFT-NLGMIGMNVGLD-RVIDQIGLANPLAQHTERLQHRIGHD 551
Qy 386 ADRESD 391
Db 552 KNLFPD 557

RESULT 7
US-10-080-170-24
; Sequence 24, Application US/10080170
; Publication No. US2004012322A9
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495, 0218
; CURRENT APPLICATION NUMBER: US/10/080, 170
; PRIOR FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270, 123
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 649
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-10-080-170-24

Query Match 7.0%; Score 140; DB 4; Length 649;
Best Local Similarity 23.5%; Pred. No. 0.0011;
Matches 114; Conservative 62; Mismatches 172; Indels 138; Gaps 25;

Qy 11 AGPVNVNMSVYLVLML--LFTWSL---ATMLP-----GTVG 41
Db 105 AALVLSVLGWLMLGTRLYAPSLQGRQAIMLPAGALVYVALPPARDPATSGLESGLV 164
Qy 42 GYPVVVWL-----LGAGAVMLASLALHETLAHVAVRAGVSVEV 84
Db 165 TYLGILMMWVWMAQPLRNSQSRFFIGALAFVAGCSVLVRPELALMGSGALIMLIMAR 224
Qy 85 TLMFGVTALGGEKTPKKAFLIAPAG---PATSLASAT---FGALAITLAGVTRPAI 138
Db 225 TCMLRALIVVAGSLPVAYLQFRMGVYGLVPGTALAKDAAGDKWSGIIYLSNPNPV 284
Qy 139 VISVAMMLATVNL--LGLFNL-----FGAPLDGCR-----L 169
Db 285 L-----WPLVLYLVGLLMLLHHRWSPFMHPLETPDGRVAVAVSPPAVVVFVPSGL 339
Qy 170 VRAYLMR-----HGSVRAGIGAAAGRVVALYLALGLAEF--VAGLVGV---W 217
Db 340 LQAFYMWIRQGDPMHGRVLLAPFLCLAPVVVLPVVISSEG--ADFSRQTGMWLAGVTSLLW 398
Qy 218 LAFIGWFIFAA---AREETRISTQQLFAGVRVADAMTAQP--HTAPGWINVEDFIORY 271
Db 399 LGVAGWSLMAANSRGMGDDATNVS---YSGIVDERRFYQAQGHANP--LTAADYL--- 449
Qy 272 VLGRHSAYPVADDSITGLVALRQLRDVAPSRRSTTSVGDIALPLHSVPTAPQPEPLT 331
Db 450 -----GYP---RMAAV--LVALNTPDGLALLPSGNYIKMDLVPMTQLSPSSGSPD 497
Qy 332 ALBERMAPLGRSRALVTESGAVVIGTVPDVARLIDVYRLAQPEPTTSPQ-----D 385
Db 498 SLVSQK----PQHTVFT-NLGMIGMNVGLD-RVIDQIGLANPLAQHTERLQHRIGHD 551
Qy 386 ADRESD 391
Db 552 KNLFPD 557
```

```
RESULT 8
US-10-468-356-24
; Sequence 24, Application US/10468356
; Publication No. US20040197896A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 05394, 0019
; CURRENT APPLICATION NUMBER: US/10/468, 356
; PRIOR FILING DATE: 2003-08-19
; PRIOR APPLICATION NUMBER: 10/080, 170
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: 60/270, 123
; NUMBER OF SEQ ID NOS: 655
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 24
; LENGTH: 649
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-10-468-356-24

Query Match 7.0%; Score 140; DB 4; Length 649;
Best Local Similarity 23.5%; Pred. No. 0.0011;
Matches 114; Conservative 62; Mismatches 172; Indels 138; Gaps 25;

Qy 11 AGPVNVNMSVYLVLML--LFTWSL---ATMLP-----GTVG 41
Db 105 AALVLSVLGWLMLGTRLYAPSLQGRQAIMLPAGALVYVALPPARDPATSGLESGLV 164
Qy 42 GYPVVVWL-----LGAGAVMLASLALHETLAHVAVRAGVSVEV 84
Db 165 TYLGILMMWVWMAQPLRNSQSRFFIGALAFVAGCSVLVRPELALMGSGALIMLIMAR 224
Qy 85 TLMFGVTALGGEKTPKKAFLIAPAG---PATSLASAT---FGALAITLAGVTRPAI 138
Db 225 TCMLRALIVVAGSLPVAYLQFRMGVYGLVPGTALAKDAAGDKWSGIIYLSNPNPV 284
Qy 139 VISVAMMLATVNL--LGLFNL-----FGAPLDGCR-----L 169
Db 285 L-----WPLVLYLVGLLMLLHHRWSPFMHPLETPDGRVAVAVSPPAVVVFVPSGL 339
Qy 170 VRAYLMR-----HGSVRAGIGAAAGRVVALYLALGLAEF--VAGLVGV---W 217
Db 340 LQAFYMWIRQGDPMHGRVLLAPFLCLAPVVVLPVVISSEG--ADFSRQTGMWLAGVTSLLW 398
Qy 218 LAFIGWFIFAA---AREETRISTQQLFAGVRVADAMTAQP--HTAPGWINVEDFIORY 271
Db 399 LGVAGWSLMAANSRGMGDDATNVS---YSGIVDERRFYQAQGHANP--LTAADYL--- 449
Qy 272 VLGRHSAYPVADDSITGLVALRQLRDVAPSRRSTTSVGDIALPLHSVPTAPQPEPLT 331
Db 450 -----GYP---RMAAV--LVALNTPDGLALLPSGNYIKMDLVPMTQLSPSSGSPD 497
Qy 386 ADRESD 391
Db 552 KNLFPD 557

RESULT 9
US-10-282-122A-50265
; Sequence 50265, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
```

```

; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykend, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034a
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 50265
; LENGTH: 706
; TYPE: PRT
; ORGANISM: Burkholderia mallei
US-10-282-122A-50265

```

```

Query Match      7.0%; Score 138.5; DB 4; Length 706;
Best Local Similarity 22.2%; Pred. No. 0.0018;
Matches 110; Conservative 48; Mismatches 159; Indels 179; Gaps 21;

Qy 7 LGRIAGFVVNV---HMSVLVILMLFTWSLATMTPGVGYPVAVVYMLGAGAVMLLASL 63
Db 121 LGRIAMVVPALAGHNV-PLIFVCGLAABAACVSAAGWRTPLRLIAGSVCMILFSA 179
Qy 64 LAHEI-----AHAVARAGVSVESVTLMLFGVTLG-----GE 98
Db 180 VTTLLAEFEQTVVGASLWASGLYOPGAAIGRDALLMLVAPLALPLVIRPLDELALGD 239
Qy 99 AKTPKAAIRIAFAGPATSLA---LSATRGALAITLAGVTPAIVISVAMWMLATVWL--- 152
Db 240 DAAAAAGVNV---DATRLAGTVAVVGPAVSAVSIAG---PLSYIG---LIAPNLIRQM 288
Qy 153 -----LGLFNLLPGAPLDGRLV----- 170
Db 289 RGAKASRIAGA---LVFLAALAGALVLVVDSAVLALGLDALTSTGVAFVGTPLMLAMIR 346
Qy 171 RAYLW---RRHGDSVRAGI-GAARAGR-----VVALVTLALGLAEFVAGLVGVWL 218
Db 347 RGAASGAAGQOPAHYRAGTRGAGRLVRAALAVLPMPAVAVLLAAGAAVLVVGASPGPMWI 406
Qy 219 AFIGFIFAAAREESTR-----ISTOOLF----- 242
Db 407 GGRFFAALAQDDIARVTLDIRAPRLCALLAGALLGASGVLMQSIIVANPLAGEBVLGV 466
Qy 243 ---AGVRVADAMTAQPHTAQGWINVED-----FIQRYVLGERHSAYPVADRDGSIQGL 292

```

```

Db 467 TQAGLATATVAVPPLAAGHVLVAASLWAGGATLVLTALNRRHRYAPLA---VALTGI 523
Qy 293 VA-----LQRLRVAPSRSTTSVGD-----IALPHSVPTAR 325
Db 524 VIGITLWTLAOWLITQSSVQPARFVVLVGGTGYRSGEAAALLPFGVLAVALAL-LAR 582
Qy 326 POEPLTALLERMAPLG 341
Db 583 PDLIALGDDEQAALAG 598

RESULT 10
US-10-156-761-9868
; Sequence 9868, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9868
; LENGTH: 570
; TYPE: PRT
; ORGANISM: Streptomyces avermiltilis
US-10-156-761-9868

```

```

Query Match      6.7%; Score 133.5; DB 4; Length 570;
Best Local Similarity 24.5%; Pred. No. 0.0038;
Matches 105; Conservative 46; Mismatches 163; Indels 115; Gaps 22;

Qy 10 IAGFVVNVHMSVLVILMLFTWSLATMTPGVGYPVAVVYMLGAGAVMLLASL 53
Db 190 LAGTV---VLWMLLPPVRRSNTIVCLGGLVPSIARYGPVAVVALLIIPVVRKPRNG 247
Qy 54 GG-----AVMLLASLAEHLAHAVVARAGVSVESVTLMLFGVTLGGEAKTPKAAE 106
Db 248 GGMTDNGRAACTLGAACAAQQLPW-----FLTPFLAGVYALRGERGPRALVV 295
Qy 107 ---RIAPGPTSLSLSTFGALAITLAGVTPAIVISVAMWMLATVWLGLFNLPGAP 163
Db 296 AVGRIGAVAACTGWLILNAFF-----VSEBPSRMSLGI-----MLPLTQSAV 336
Qy 164 LDGRLVAVYMRBHGDSVRAGI-GAARAGR---VVALVTLALGLAEFVAGLVGV- -WLAFI 221
Db 337 IHGGVLGASLYYDGGSLAMV---SHASTLAAVLALVLFPKRRLGPRATVLPKCF- 393
Qy 222 GWFIFAAAREESTRISTOOLFAGVRVADAMTAQPH-APGWINVEDFIQRYVLGERHS- 278
Db 394 ---FLATRSQD---GYLLMTPLMLAAAVTAPPSAFASAW-----QPLRPFGAHRRRA 440
Qy 279 -----AYPADRDGSIITGLVALROLQADVAPSRSTTSVGDIALPLHLSV---PVARP 326
Db 441 ARVGVALLLALLPLSAAAYATGAPPLRM---EVALVRRGVTGTGLTLRVYNTGGTTMRP 498
Qy 327 QEPLTAL---LBR---MAPLGPBSRALVTEGSAVVGIVTPSDVARLIDVYRLAQPEP----- 377
Db 499 HFTVITNQGMRKYSVAGAPAS---VPAHGSARYELBLPPDG-----RFLPLPFGVRI 548
Qy 378 ---TFTTSPO 384

```

Db 549 LRASASP 557

RESULT 11

US-10-493-462-25

Sequence 25, Application US/10493462

Publication No. US20040253711A1

GENERAL INFORMATION:

APPLICANT: James, Brian William

APPLICANT: March, Philip

APPLICANT: Hampshire, Tobias

TITLE OF INVENTION: Mycobacterial Genes Down-Regulated During Latency

FILE REFERENCE: 1581.1040000

CURRENT APPLICATION NUMBER: US/10/493,462

PRIOR FILING DATE: 2004-04-23

PRIOR APPLICATION NUMBER: PCT/GB02/04718

PRIOR FILING DATE: 2002-10-21

PRIOR APPLICATION NUMBER: GB 0125535.5

NUMBER OF SEQ ID NOS: 46

SOFTWARE: PatentIn version 3.1

SEQ ID NO 25

LENGTH: 944

TYPE: PRT

ORGANISM: Mycobacterium tuberculosis

US-10-493-462-25

Query Match 6.7%; Score 133; DB 5; Length 944;

Best Local Similarity 24.4%; Pred. No. 0.0083;

Matches 94; Conservative 43; Mismatches 152; Indels 96; Gaps 18;

76 RAGVSVESVTMLFGVTAL-----GGEAKTPKAARFIAFAGPATSLATPGALAIT 129

529 RAIFPRKQITV-LVGGTALBELDSIHGLPAKMP---LWVILLTTIYLMFLAPGSVLP 584

130 LAGVTPAIVSVAMWLTATVNLILFLF-----NLLP---GAPLDG----- 166

585 IKATLMSALTLG---STMGILTWIFVDGHFSKMLNTPPLTAPVIGLIYALVFGSLT 639

167 -----GLRVAYIMRHGDSVRAG--IGABAGRV--ALVLTALGLAEFVAGLVGG 215

640 DYEYFLVRMVEA---RRGGSTQEARIGTATRTITTAALIVAVVAGAFVSSDLVMM 696

216 VMLAFIGWFIAPAAAREETRISTQOLFAGVRVADMTAOPHTAPGW-----IN 263

697 KYLAN---GLMALLLDATVV---RMFLVPSVMKLGDDCMWAPRMARLQTRIGLGEIH 750

264 VEDFIQRYVLGERSAVYVADRDSITGLVALRQLRDVAPESSRTTSVGDIALPLHSVPT 323

751 LPDERKRVNSGRPARPVPT-----AGLVAAAGADPRPHDPHTPLAESPRPARSSPA 804

324 ARQPPRLALERMALGPRSRALVTESAVVGIYTPSDVARLILVYRLAQ-----PEPT 378

805 SSPE--LTPALEATYAAPAPSGASTR--MQIGSSTEPTTKLAAAGRSVSPASTPPT 860

379 FT-----TSPQADARESDA 392

861 PTPPSAPSAAGOTRAMPLAANRTDA 885

RESULT 12

US-10-205-032-14

Sequence 14, Application US/10205032

Publication No. US20030113874A1

GENERAL INFORMATION:

APPLICANT: Farnet, Chris

APPLICANT: Yang, Xianhu

APPLICANT: Staffa, Alfredo

TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN

FILE REFERENCE: 3016-2US

CURRENT APPLICATION NUMBER: US/10/205,032

CURRENT FILING DATE: 2002-07-26

NUMBER OF SEQ ID NOS: 39

SEQ ID NO 14

LENGTH: 3362

TYPE: PRT

ORGANISM: Streptomyces halstedii sp. NC-34

US-10-378-083-6

Query Match 6.6%; Score 130.5; DB 4; Length 3362;

Best Local Similarity 24.1%; Pred. No. 0.079;

Matches 87; Conservative 52; Mismatches 153; Indels 69; Gaps 14;

47 VYMLIGAGAVMLASLAEHLAAVAVRAGVS-----VESYTLMLFGVTA 94

416 VDW---SAGAVELTE--AREWPEYGRPRAGVSPFGSGTMHVIIEEASEPESAVEP 470

95 LGGEAKTP-----KAAFRIFAGPATSLATPGALAITLAGVTPAIVT 140

SOFTWARE: PatentIn version 3.0

SEQ ID NO 14

LENGTH: 3745

TYPE: PRT

ORGANISM: micromonospora carbonacea subsp. aurantiaca

US-10-205-032-14

Query Match 6.6%; Score 132; DB 4; Length 3745;

Best Local Similarity 25.2%; Pred. No. 0.066;

Matches 107; Conservative 40; Mismatches 174; Indels 104; Gaps 20;

31 SLATMLPGTYGTYAVYTYMLLGGAGVMLASLL-AHELHAAVAVRAGVS----- 83

1063 SLTLAAGADARHRAVP---LGLTASLALQALADACTORLMAVTRGAAVASSGEVDA 1119

84 --VTLMFGVTVL-----GGEAKTPKAARFIAFAGPATSLATPGALAITLAG--- 132

1120 GQAQVWGLGRVYALDELDRKGGVLDLPALGERAFALADLVGGSGNEDQVAVRAGSVYG 1179

133 ---VTPAIVSVAMWLTATVNLILGFLNLPFGAPLDGRLVRAVYLMRHGDSVRAIGAA 189

1180 RRLVRSRAVTVSGDMPARGTILVVG--DTGVALALAGRL-----GDGAA 1223

190 R---AGRVVALVTLALGLAEFVA-----GLVG-----GYMLAFIG 222

1224 HVVLAGEPAAASTVGLTGADRVALLIDCDPSDRDALGLAYRPTTVVAPPAVALTALA 1283

223 W-----FIPAAAREETRISTQOLFAGVRV-ADAMTAOPHTAPGMINVEDPIGRVYGER 276

1284 ETTEDETVAAVAAKTTTAAVHLDALAELELDAFVVFSSVSGTW-----CGAG 1332

277 HSAYPVADRDSITGLVALRQLRDV-----AESRRTSVG---DIALPLHSVPTAR 325

1333 HGCV--AAGTARLDALVEERRARGLPATAIAMTPWADATTAAGQAPDASAGCHE-PDTR 1389

326 PQEPLTALLER--MAPGPRSRALVTEGSAV--VGITYPBDV--ARLIDVYRLAQPEPTF 379

1390 AGGDRELRIRGGLTPLDPGALDVLRGAVARGELTVADVAVRAGVASTAARPTTLF 1449

380 TTSPO 384

1450 DELPE 1454

RESULT 13

US-10-378-083-6

Sequence 6, Application US/10378083

Publication No. US20040053274A1

GENERAL INFORMATION:

APPLICANT: President of Tokyo Institute of Technology

TITLE OF INVENTION: A gene cluster of vicenistatin biosynthase, a vicenistamine

TITLE OF INVENTION: glycosyltransferase polypeptide, and a gene encoding the

TITLE OF INVENTION: polypeptide

FILE REFERENCE:

CURRENT APPLICATION NUMBER: US/10/378,083

CURRENT FILING DATE: 2003-03-04

NUMBER OF SEQ ID NOS: 26

SOFTWARE: PatentIn version 3.1

SEQ ID NO 6

LENGTH: 3362

TYPE: PRT

ORGANISM: Streptomyces halstedii sp. NC-34

US-10-378-083-6

Query Match 6.6%; Score 130.5; DB 4; Length 3362;

Best Local Similarity 24.1%; Pred. No. 0.079;

Matches 87; Conservative 52; Mismatches 153; Indels 69; Gaps 14;

47 VYMLIGAGAVMLASLAEHLAAVAVRAGVS-----VESYTLMLFGVTA 94

416 VDW---SAGAVELTE--AREWPEYGRPRAGVSPFGSGTMHVIIEEASEPESAVEP 470

95 LGGEAKTP-----KAAFRIFAGPATSLATPGALAITLAGVTPAIVT 140

```

Db      471 IAGSGVTPPWVLSASADALRCQABRLISFVSAAQDVSVVDVAYSLSGVSRAGLEHRRGVV 530
Qy      141 SVAMWMLATVNLIGLFNLLPGAPLDC-----GRLVYATLMRRHGDVVRAGIGAARAG 192
Db      531 GE-----SRRLAALLESLSASGVSPCVTGRVAEGRL--AFLPTGOG-AQRGMKRELA- 582
Qy      193 RVVALVLIALGLAEFVAGLVGVWMLAFIGWFIFAAAREEETRISTOOLPAGVRVADAMT 252
Db      583 --AAPPFLFAASLE--TCGLTERAGVA--VREVLFAEESGSAEAAALLTRVYAOALFAVEV 637
Qy      253 AQPHTAPGHNVEDFIQRVYGERHSAY--PVADDGSLTGVALRQLRDVAPSRRTS 310
Db      638 ALFRLVSEFVGVDPFVAGHSGVEIAAAHYAGVSEDAVSLVAABGRMLDALPEGANVA 697
Qy      311 VGDIALPLHSVPTAPRPOEPLTALLERMAPLGRSRALVTESGAVVGIYTPSDVARLIDVY 370
Db      698 V-----QATEEDVLLALLEGVE---DASIAIANGPDAVVSSTGEAVARVVDVL 742
Qy      371 R 371
Db      743 R 743

```

```

RESULT 14
US-10-156-761-10433
; Sequence 10433, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10433
; LENGTH: 4685
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10433

```

```

Query Match      6.6%; Score 130.5; DB 4; Length 4685;
Best Local Similarity 23.6%; Pred. No. 0.12;
Matches 118; Conservative 47; Mismatches 168; Indels 167; Gaps 26;

Qy      7 LGRIAGPVNVVMSVLTVMILFTMSLATVLP-----TVGGVPVNVV----- 49
Db      997 VGRISG-----TWLVVT--PGGAPVEEVRALAVGAEAIVLAMEBE 1038
Qy      50 -----LLGAGAVMLASLLAHELAHAAVVA-----RRAGVSESVTM-LFGVTA 94
Db      1039 GREALAGRLTGA-GGVSGVSLIGWDEEAAPVATVTLVQALDAGVEA-PLVVLITQGAAS 1096
Qy      95 LGGEAKTRKAAPRIAPAPATSLASATPGALATITLACVTRPATIYISVAMWMLATVNLIG 154
Db      1097 VGAEEVHPVQTVQVVALQVAGLEQPGSWGGL-VDPGVWDERVASGLTAVIAA----- 1149
Qy      155 LFNLLPGAPLD-----GRLVRAVYL-----WRRHGSV-----RAGIGA--- 188
Db      1150 -----GEGEDOVAVRSSGAYARRLVRAPLGANPAVADNMHNGVTLITGGGGIGAHIA 1203
Qy      189 ---ARAGRVVALVLI-----ALGLAEFVA--GGLVGVWMLAFIGWFIFAAAREETRIS 237

```

```

Db      1204 RMLAKEG-AERLLVSRSGEQAEGAEIATELGL--GAETVF-----AACDVSRDA 1253
Qy      238 TOOLFAGRVADAMTAQPHTA--PGMINV-----EDFIQRYVLGERH----- 277
Db      1254 LAHVINGIPAHHPILTAHVHTGISYABELATATPEHPFDVLSARVLGARHIDELTAELGV 1313
Qy      278 --SAYPV-----ADROGSITGLVALROLRDVADS-----RSTSVG 312
Db      1314 ELEAFVFPSSGAAVVWGSNGNANAAAGYLDGLVRRRARAGVAGTSVWGGWQATVAMAVG 1373
Qy      313 DIALPLHSVPTAPRPOEPLTALLERMAPLGRSRALVTESGAVVGIYTPSDVARLIDVRL 372
Db      1374 DTAEOI-SRRGVRLLDPVLAIV-----QALROVLEODEVSVTYTDMDMALFTPGYAM 1423
Qy      373 AQPHTFTTSPQDADRFSDA 392
Db      1424 ARRRRLIEDIPEAGALSQA 1443

```

```

RESULT 15
US-10-437-963-177542
; Sequence 177542, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; PRIOR FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 177542
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_75183C.1.pep
US-10-437-963-177542

```

```

Query Match      6.4%; Score 128; DB 4; Length 487;
Best Local Similarity 22.7%; Pred. No. 0.0097;
Matches 54; Conservative 33; Mismatches 93; Indels 58; Gaps 9;

Qy      52 GAGGAVMLASLLAHELAHAAVARAGVS-----VESVTMLPGCVTALGGEAKTPKAA 105
Db      241 GYVGLVTAALITGVHEIHAIHAARDTGKIAVPYFVPSWQIGSGALTRINIVYANREDL 300
Qy      106 FRIAPAGPATSLA-----SATPGALATITLGA---V 133
Db      301 LKVAAGPLAFAFSGLFVLLIGFTLPSPDGLVIDPAVPHESFLVGLATLIGDLMKE 360
Qy      134 RTPAIVISVAMWMLATVNLIGLFNLLPGAPLDGRLVRAVYIMRRHGDVVRAGIGAARAG 193
Db      361 GTKLINPLVIM-AWAGILLAINSIPAGEIDGGRIIFA-WM-----GRKISSR 407
Qy      194 VVALVLIALGLAEFVAGLVGVWMLAFIGWFI-----FAAAREETRISTOOLPAGVRY 247
Db      408 ISSIALIGLISALFND--VAFYVVL-LFPLQKGPISPLSEETPEBNNTISIGVAL 462

```

Search completed: March 23, 2006, 06:28:17  
Job time : 84.6454 secs



**This Page Blank (usp)**



GENERAL INFORMATION:  
APPLICANT: Bailey, Richard B.  
APPLICANT: Blomquist, Paul  
APPLICANT: Doten, Reed  
APPLICANT: Driggers, Edward M.  
APPLICANT: Madden, Kevin T.  
APPLICANT: O'Toole, George  
APPLICANT: O'Toole, Jessica  
APPLICANT: Trueheart, Joshua  
APPLICANT: Walbridge, Michael J.  
APPLICANT: Yorgey, Peter S.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID  
TITLE OF INVENTION: PRODUCTION  
FILE REFERENCE: 14184-030001  
CURRENT APPLICATION NUMBER: US/10/858,730  
PRIOR FILING DATE: 2004-06-01  
PRIOR APPLICATION NUMBER: US 60/475,000  
PRIOR FILING DATE: 2003-05-30  
PRIOR APPLICATION NUMBER: US 60/551,860  
PRIOR FILING DATE: 2004-03-10  
NUMBER OF SEQ ID NOS: 364  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 104  
LENGTH: 529  
TYPE: PRT  
ORGANISM: Mycobacterium tuberculosis  
US-10-858-730-104

Query Match 6.0%; Score 119; DB 6; Length 529;

Best Local Similarity 27.1%; Pred. No. 0.098; Mismatches 143; Indels 108; Gaps 21;  
Matches 105; Conservative 32;

11 AGPVNVMSVYLMLFTWSLATMLPGTVG-----GYPAVVYMLGAGCA--VML 59  
166 AGFALGVA-MLGGTWL-TCVLAATSGVIDRLRLNRIGTLPFGQVFGAGIATLVAV 223  
60 LASLAEHLAAVAVARRAGVSVETLMLFGVTLGCG--EAKTPKAAFIAPGAPATSL 117  
224 AAYIAGODPTALVA--TGIIV-----LISGMTLVSMODATVGYMLTALARIGDALFL 275  
118 ALSATFGALITTLGAVTPAIVISV-----AMWLATVNLGLFNLLPGAPLDGGRL-VRA 172  
276 TAGIVGVL-ISLRGVTNAGIQIELHVDATTTLATPGMPPLIVAVSGALSGVCLTIAS 334  
173 YLMRRHDSVRAGIGAPARAGVVALVLIAGLAEFVAGLVGVTWLAFIGFIAPAREE 232  
335 YAPLR--SVATNAGLSAGLA---ELVILGLGAGF--GRVATWTATAIIGVFLA----- 380  
233 ETRISTQQLFAGVAVADAM-TAOPHTAPGMINVEDFIQRYVLGERHSAYPVADRDSITG 291  
381 -TLISIR-----RQAPALVTATGIMP-----MLPG 405  
292 LVALRQLRDVAPSRSTTSVGDIALPLHSVPTAPQEPPLTALERMALGPRSRALVTEG 351  
406 LAVRAVAFAPFA-----VND-----TPDGLTOLLEAA-----TALALGS 440  
352 SAVVGIVTPSDV---ARLIDVYRLAQP 375  
441 GVVLGEFLASPLRYGAGRIGDLFRIEGP 468

## RESULT 3

US-10-858-730-105  
Sequence 105, Application US/10858730  
Publication No. US20050255566A1  
GENERAL INFORMATION:  
APPLICANT: Bailey, Richard B.  
APPLICANT: Blomquist, Paul  
APPLICANT: Doten, Reed  
APPLICANT: Driggers, Edward M.  
APPLICANT: Madden, Kevin T.  
APPLICANT: O'Toole, Jessica  
APPLICANT: O'Toole, George

APPLICANT: Trueheart, Joshua  
APPLICANT: Walbridge, Michael J.  
APPLICANT: Yorgey, Peter S.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID  
TITLE OF INVENTION: PRODUCTION  
FILE REFERENCE: 14184-030001  
CURRENT APPLICATION NUMBER: US/10/858,730  
CURRENT FILING DATE: 2004-06-01  
PRIOR APPLICATION NUMBER: US 60/475,000  
PRIOR FILING DATE: 2003-05-30  
PRIOR APPLICATION NUMBER: US 60/551,860  
PRIOR FILING DATE: 2004-03-10  
NUMBER OF SEQ ID NOS: 364  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 105  
LENGTH: 529  
TYPE: PRT  
ORGANISM: Mycobacterium tuberculosis  
US-10-858-730-105

Query Match 6.0%; Score 119; DB 6; Length 529;

Best Local Similarity 27.1%; Pred. No. 0.098; Mismatches 143; Indels 108; Gaps 21;  
Matches 105; Conservative 32;

11 AGPVNVMSVYLMLFTWSLATMLPGTVG-----GYPAVVYMLGAGCA--VML 59  
166 AGFALGVA-MLGGTWL-TCVLAATSGVIDRLRLNRIGTLPFGQVFGAGIATLVAV 223  
60 LASLAEHLAAVAVARRAGVSVETLMLFGVTLGCG--EAKTPKAAFIAPGAPATSL 117  
224 AAYIAGODPTALVA--TGIIV-----LISGMTLVSMODATVGYMLTALARIGDALFL 275  
118 ALSATFGALITTLGAVTPAIVISV-----AMWLATVNLGLFNLLPGAPLDGGRL-VRA 172  
276 TAGIVGVL-ISLRGVTNAGIQIELHVDATTTLATPGMPPLIVAVSGALSGVCLTIAS 334  
173 YLMRRHDSVRAGIGAPARAGVVALVLIAGLAEFVAGLVGVTWLAFIGFIAPAREE 232  
335 YAPLR--SVATNAGLSAGLA---ELVILGLGAGF--GRVATWTATAIIGVFLA----- 380  
233 ETRISTQQLFAGVAVADAM-TAOPHTAPGMINVEDFIQRYVLGERHSAYPVADRDSITG 291  
381 -TLISIR-----RQAPALVTATGIMP-----MLPG 405  
292 LVALRQLRDVAPSRSTTSVGDIALPLHSVPTAPQEPPLTALERMALGPRSRALVTEG 351  
406 LAVRAVAFAPFA-----VND-----TPDGLTOLLEAA-----TALALGS 440  
352 SAVVGIVTPSDV---ARLIDVYRLAQP 375  
441 GVVLGEFLASPLRYGAGRIGDLFRIEGP 468

## RESULT 4

US-11-205-109-15  
Sequence 15, Application US/11205109  
Publication No. US20050287641A1  
GENERAL INFORMATION:  
APPLICANT: Farnet, Chris  
APPLICANT: Zazopoulos, Emmanuel  
APPLICANT: Stafila, Alfredo  
TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS  
FILE REFERENCE: 3002-2US  
CURRENT APPLICATION NUMBER: US/11/205,109  
CURRENT FILING DATE: 2005-08-17  
PRIOR APPLICATION NUMBER: US/09/976,059  
PRIOR FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: US 60/239,924  
PRIOR FILING DATE: 2000-10-13  
NUMBER OF SEQ ID NOS: 46  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 15  
LENGTH: 8695

```
; TYPE: PRT
; ORGANISM: Actinoplanes sp.
US-11-205-109-15

Query Match
  6.0%; Score 118.5; DB 7; Length 8695;
Beet Local Similarity 26.2%; Pred. No. 3.1; Indels 143; Gaps 24;
Matches 104; Conservative 32; Mismatches 118; Indels 143; Gaps 24;

QY 50 LLGAGAVMLLASLHAHELAHAHVARRAGVSVEVT--LW---LFGVTA-----LGG 97
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 4899 LVGAGAVVAEDAVRDPAVAALLDRHAVTVQATPMLQMALLAGHDAVADVLLVGG 4958
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 98 EAKTPKAAFRIFAPGPA-----TSLASATFGALATLAG---VRTP-----AIVISV 142
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 4959 EALPPALAGRMAAARGVTNLVGPTEVTWATVADLGASPAAPVPIGTPLRTRAFVLLD 5018
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 143 AMWLATVNLGLFNLPGAP---LDGRLVRAVIMRHGDSVAGIGAAAGAVMLV 198
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 5019 A-----LRPVPGVPEBELYIAGDQLARGY---HG---RAGLTAER-----FV 5054
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 199 LIALGLAE--FVAGGLV---GVMILAFIGMFIFAAREEETRISTQQLFAGVRVADAMT 252
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 5055 ADPFRGERMYRTGDRVKTREGS--LEFIG-----RVDDQVKIRGFRIE----- 5097
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 253 AQPHTAPGWINVEDFIQRVYIGERSAY---PVADROGSITGLVALRQLRDVARSRSST 308
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 5098 -----LGEVEAALLAFGPVA-----RAAAAVRE--DVPGDRRLV 5129
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 309 TSVGDI-----LPIHSVPR--RROEPTTA--LIERMAPLGP----- 342
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 5130 GVVPAAGEPEPDPAVRAHVAQAQLPAYVVPFAVVVLPPLPLTANGKLDKRLAPADYGA 5189
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 343 --RSRALVTESGAVVGIYTP---SDVARIIDVYRL 372
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 5190 ASAGRAPADEREALICAVPAETLGTVDVAADDFPAL 5226
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

RESULT 5
US-11-205-109-23
; Sequence 23, Application US/11205109
; Publication NO. US20050287641A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
; FILE REFERENCE: 3002-205
; CURRENT APPLICATION NUMBER: US/11/205,109
; CURRENT FILING DATE: 2005-08-17
; PRIOR APPLICATION NUMBER: US/09/976,059
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/239,924
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Actinoplanes sp.
US-11-205-109-23

Query Match
  5.8%; Score 115; DB 7; Length 403;
Beet Local Similarity 20.5%; Pred. No. 0.14;
Matches 80; Conservative 39; Mismatches 135; Indels 136; Gaps 15;
```

```
QY 149 VNLGLFNLPGAPLOGRLVRAVIMRRHDSVRAGIGAAAGRVNVALIALGLAEFV 208
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 120 -----VLTGSVRAAPTIIPAF-----AVGGAALG--VVALSQVAATW-- 156
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 209 AGGLVGVMLAFIGNF-----IFAAREBETRISTQQLFAGVRVADAMTAQHT 257
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 157 AGPAPVTWMLITLWLGSTANGLSIRYDGAARANAERIROBERIELAR----- 204
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 258 APGWINVEDFIQRVYLG---ERHSAYPVADRDG-----SITGLVALROLDVAPS 304
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 205 -----ELHDVVAHHITGMILDTQAQAVLARDAGVPERLAVIETPAETALAAARVYGL 259
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 305 RRTSTVSVDIALPLHSVTPARPOEPLTALLERMAPLGRSALVTESGAVVGIYTPSYA 364
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 260 LRADAD-----GPPSAPEPEELSTLVERFSROGPFVRLTTPDKMKOMPVEVTTYV 310
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 365 RL-----DVRILA--OPEPTTSPQAD 387
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 311 RIVREALTNVAHAPHAPNVTVTVTEQAD 340
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

RESULT 6
US-11-075-185-5
; Sequence 5, Application US/11075185
; Publication NO. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; LENGTH: 3655
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-11-075-185-5

Query Match
  5.7%; Score 112.5; DB 7; Length 3655;
Beet Local Similarity 21.9%; Pred. No. 3.1; Indels 205; Gaps 26;
Matches 110; Conservative 44; Mismatches 144; Indels 205; Gaps 26;

QY 51 LGAGAVMLLAS-----LHAHELAHAVARRAGVSVEVTMLFGVYTAGGE 98
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 1423 LDPGTVLITGTGTELGQVRAHILVAHGVNHLVLTSRGDADD-----AALLVDE 1474
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 99 AKTPKAA-FRIAPGATSLASATFGAL-----AITLAVRTPAIVISV-AMWLAT 148
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 1475 LRAAGAAATVDVAACDVADGALGAVIAIIPAHPDLTAVHMAVGLDVIYTKLSABQLAR 1534
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 149 VNLGLFNLPGAPLOGRLVRAVIMRRHDSVRAGIGAAARVVALVLI----- 200
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 1535 V-----LRP--KIDGG-----WHL-----AAATGRHLAEFVFPSSAAGTLG 1569
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 201 ALGLAEFVAG-----GLVGVW-----LAFIGMFIFAA--AREE 232
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 1570 SAGQANVYAAAFDALLAAQLRARGVPRMSLAKMFWEAGAGMTAHLG-----AADLARLR 1625
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 233 ETRISTQQLFAGVRVADAMTAQHTA--PGWINVEDFIQR----- 270
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 1626 RQGIAPILAAQGMQLDRALARPALVLPALDL--SALQRAASDAGQVALLRGLVPRPA 1684
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 271 -----YVGERSAVPVADROGSITGLV-----ALROLADVA--DSRRSTT 309
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 1685 GRRAASPAATAAGAAALRLARLSLPEARAGALLLELVAAEAAVYQLAGPAQVPAKPLK 1744
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
```

```
Oy 310 SVG-----DIALPL-----HSVTPARPQEPYTLALERMALPG----- 341
;
Db 1745 ELGLSTLAVBELRNKLGARAEATLPAFDHPTFPAIADLLQRAFSLELAAGTTRAQA 1804
Oy 342 PRSRALVTEGSAVV-----GIYTPSDVARI-----PYVRAJOEP 377
;
Db 1805 PRARAHDEPILNIVSMACRLPGSVDTFPAIMQLBLEGDAIGPFPGMGVYAGLYDIP 1864
Oy 378 TTTTSP-----ODADRF 389
;
Db 1865 ---DAFGKSVTNLGGFLYDADRF 1884

RESULT 7
US-11-087-099-10617
; Sequence 10617, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 10617
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Torenia hybrida
US-11-087-099-10617

Query Match 5.6%; Score 112; DB 7; Length 511;
Best Local Similarity 24.9%; Pred. No. 0.32;
Matches 104; Conservative 48; Mismatches 146; Indels 120; Gaps 22;

Oy 21 VLVIIMLFVSLATM-----LPGTVGYPVVVYVWLLGAGAVMLLASLAHELAHAV 73
;
Db 19 VLVIITLVSKSLITIVSRKRLPPGPTGFPVY-----GALPILGSM--PHYALAKM 68
Oy 74 ARRAG--VSBSVTMLPGCVTALGGEA--KTPKAAF--RIAFAGPATSLASATFGALA 127
;
Db 69 AKKGPVWYVLTGTVGAVASTPSARAFKLTLDNFSNRPNAG--ATHLA---YGAQD 123
Oy 128 ITLA--GVRTPAIVSVAMWLATVNLGLFVL--LPGAPLDGGLVYAYLMRRRGGSVR 183
;
Db 124 WVFAYGRR-----W---RLRLRLSNLHMLGKALDGMANVNA--SELGHMLE 166
Oy 184 AGIGARAGRYVALVIALGLAEFVAGLVGVMIAFGMIFFAAREERTISTQQLFA 243
;
Db 167 AMNRASRGEVV---VPEMLVYAMANNIGV---ILSRVFTVKGEE----- 208
Oy 244 GVRVADAMTAQPHTPAPGWINVEDFIQRYVLGERHSAYPV---ADRDGSIQGLVALROLRD 300
;
Db 209 -VNERKQVWVVELMTSAGYFNIGDV-----PLAMMDLQGIERGMKALHKKFD 255
Oy 301 -----VAPS-----RSTTSVGDIALPLHSVTPARQOE 328
;
Db 256 KLITRMLQHEASAHKPDFLDIAIASJDDYSSEERLSTSNIKALLNLPLFAGDTSSS 315
Oy 329 PLTLALERMALPGRSRALVTEGSAVVG---LVTPSDVARIADYVRLQAPPTTSP 383
;
Db 316 TTEWALSEMLKKKGLKKAQOEEMDRVGRERRLVESDIEKL--GYLKAIKCTETPKHP 371

RESULT 8
US-11-143-980-61
; Sequence 61, Application US/11143980
; Publication No. US20050272133A1
; GENERAL INFORMATION:
; APPLICANT: He, Min
; APPLICANT: Hucul, John
; APPLICANT: Hallel, Bradley A.
; APPLICANT: Wagenaar, Melissa M.
; APPLICANT: Graziani, Edmund
```

```
; APPLICANT: Summers, Mia
; APPLICANT: Kulowaki, Kerry
; APPLICANT: Pong, Kevin
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Production of a Complex
; FILE REFERENCE: AM-101426US
; CURRENT APPLICATION NUMBER: US/11/143,980
; CURRENT FILING DATE: 2005-06-03
; PRIOR APPLICATION NUMBER: US 60/664,483
; PRIOR FILING DATE: 2005-03-23
; PRIOR APPLICATION NUMBER: US 60/576,895
; PRIOR FILING DATE: 2004-06-03
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 61
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Streptomyces sp.
US-11-143-980-61

Query Match 5.6%; Score 110.5; DB 7; Length 497;
Best Local Similarity 24.2%; Pred. No. 0.41;
Matches 107; Conservative 40; Mismatches 144; Indels 151; Gaps 22;

Oy 1 MRDAIPGRIRAGFVNNHMSVTVLMLFTWSLATMLPGTVGYPAV-----VY 48
;
Db 130 VRDHPSPRPG-----LIRMQTMAAMWMLIGVGG--PLGGLLADHIGWRMAF 177
Oy 49 WL---LG-AGGAVMLLA-----SLAHELAHAV--VARRAGSVES 83
;
Db 178 WLNPLGLAAGAVVVLPLPRRRPATPPSGRLDVAGIILLAAGLALTLGSLKGNATAGH 237
Oy 84 VTLW-----LFGVYAL-----GGEAKTPRAARIFAFAGPATSLASATFGALAIT 129
;
Db 238 APSWTDPAVIGCLIGLALTLTLIPVERRAAVPLPLFRHRYTALITAGFFQVAAA 297
Oy 130 LAGRTPAIVSVAMWLATVNLGLFNLPGAPLDGGLVYATLMRR--RGDSV----- 182
;
Db 298 PVGIFLPEYFOHIGHSATASGLL--LPLILGML--SNRLTAALVLSGHVKKPVLIGAG 355
Oy 183 ---RAGIGARAGRYVALVIALGLAEFVAGLVGVMIAFGMIFFAAREERTIS 237
;
Db 356 LITAGTAFAVLRITTEPLALTSVLLLVGLGAGPMGGLTIA-----TQSA 401
Oy 238 TQQLFAGRYVA--DAMTAQPHTPAPGWINVEDFIQRYVLGERHSAYPADRDGS-----I 289
;
Db 402 VPRADMGTATGSLTQGLGAVGLAQAQSLI-----GHSQAAPATATAGSTVSWSGA 456
Oy 290 TGLVALQL---RVASRSRSTTSVGDIALPLHSVPTA--RQPEPLTALERMALGPRS 344
;
Db 457 AGLTALGALLMRDI-----SIATAGKRPAP----- 483
Oy 345 RALVTEGSAVVGIYTPSDVARL 366
;
Db 484 ----TSGTAV-----PAKADRL 496

RESULT 9
US-10-467-657-2362
; Sequence 2362, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASTIGNANI Vega
; APPLICANT: MONACT Elisabeta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
```

NUMBER OF SEQ ID NOS: 9218  
SOFTWARE: SeqWin99, version 1.04  
SEQ ID NO: 2362  
LENGTH: 488  
TYPE: PRT  
ORGANISM: Neisseria gonorrhoeae  
US-10-467-657-2362

Query Match 5.5%; Score 109; DB 6; Length 488;  
Best Local Similarity 24.7%; Pred. No. 0.52;  
Matches 78; Conservative 41; Mismatches 131; Indels 66; Gaps 16;

QY 30 WSLATMLPCTVG-----YPAV---VYMLLGGAGVMT----- 59  
DB 60 WA-ASILLPASAELPMWKKPYSAVNSGDTAVWMTAAVLMTLPGALFYGQWVRKK 118  
QY 60 LASLAHBLAHVAVRAGVSVEVTLMFGVTLGGEAKT-----PKAFRIAPAGP 113  
DB 119 LLSLTHHSGSITLTVGILWAV--GYSLATPGNAFIGLGRVPLSGMOTDAARMLTSP 177  
QY 114 ATSLASATFGALATTLACVTPAIVISVAMWL--ATVNLGLFNLL--PGA--PLDG 166  
DB 178 NAFPTPEPVMFQMTFALITGAFERMKYSAMWLFGIWFLLVYVPGAHVWVG 237  
QY 167 GLVPAVYMRHSGSVRAGIGARAGRVVALV---IALGLAEF---VAGLVGGWILA 219  
DB 238 GEMSKGGLVDYAGCTV--VHINAGIAGLVAALVIGRRIGYGRAMPPEHMAATLIGAMLM 296  
QY 220 FFGWTFFAAREERISTQQLFAGVRVADMTAGHTAPGHTINVEDFQIRVLEGRHA 279  
DB 297 F-GWGFNG-----SALAADAAGAMAVTGVSAVFGAAGWLACER-----IAGHKPSA 345  
QY 280 YPAADRGSIITGLVAL 295  
DB 346 LGLA--SGAVSGLVGI 359

RESULT 10  
US-11-087-099-1493  
Sequence 1493, Application US/11087099  
Publication No. US20060041961A1  
GENERAL INFORMATION:  
APPLICANT: Adac, Mark S. et al.  
TITLE OF INVENTION: Genes and Uses for Plant Improvement  
FILE REFERENCE: 38-21(53450)B EP  
CURRENT APPLICATION NUMBER: US/11/087,099  
CURRENT FILING DATE: 2005-03-22  
NUMBER OF SEQ ID NOS: 12464  
SEQ ID NO 1493  
LENGTH: 514  
TYPE: PRT  
ORGANISM: Thermobifida fusca  
US-11-087-099-1493

Query Match 5.5%; Score 109; DB 7; Length 514;  
Best Local Similarity 21.2%; Pred. No. 0.55;  
Matches 87; Conservative 46; Mismatches 126; Indels 152; Gaps 18;

QY 44 PAVVYVWLGGAGVAMLSLAHBLAHVAVRAGVSVEVTLMFG----- 90  
DB 73 PAIVISFL--AAGAVCLAMCYAEFASIVPVAGSAVTFGATLGEFVAMWIGMDILBEFT 131  
QY 91 ---GVTALGGEAKTPKAFRIAPAGPATSLASATFGALATTLACVTPAIVISVAMWLA 147  
DB 132 LAASVSVGMS-----EYAGDLALPTTVTLAGLPVNLGAV----- 167  
QY 148 TVNLLGLFNLLPGAFLDGRVRAVYMRHSGSVRAGIGARAGRVVALVIALGLA-- 205  
DB 168 AIAVLLGLVGM--GATLSG-----RVTAVVVALKVG--IVLFIIVAGAAV 210  
QY 206 -----EFVAGLVGG---VWLATIGWFI 225  
DB 211 DEANWTPPIPAQPAADTTGVSDATLTHTVLFGLBEFTSPGWPVVAAASVFEFSFG--FD 269

QY 226 FAAREESTRISTQQLFAGV-----RVADAMTA-QPH-----TAP----- 259  
DB 270 IVATTAESTRNPBRDIPVGIFGSLIITVLYMAVAAVTGMKRPYELNVTAPLSDAFRSV 329  
QY 260 -----GWINVEDFIQRYVILGERHSAVPADDSITGLVALRQLRDVAPSRRS 307  
DB 330 GADWAAATLISLUGIGITTVTLVLMGQARVAFWS--RDGLLPRLSIRHPRFGTPYTT 388  
QY 308 TTSVGDIALPLHSVPTARPOEPLTALERMAPLGRSPALVTGSAVVGIV 358  
DB 369 LITTAVALLGIVP-----ISTLEEMNIGTLFAFVVS-----YGVV 427

RESULT 11  
US-11-205-109-30  
Sequence 30, Application US/11205109  
Publication No. US20050287641A1  
GENERAL INFORMATION:  
APPLICANT: Farnet, Chris  
APPLICANT: Zazopoulos, Emmanuel  
TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS  
FILE REFERENCE: 3002-2US  
CURRENT APPLICATION NUMBER: US/11/205,109  
CURRENT FILING DATE: 2005-08-17  
PRIOR APPLICATION NUMBER: US/09/976,059  
PRIOR FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: US 60/239,924  
PRIOR FILING DATE: 2000-10-13  
NUMBER OF SEQ ID NOS: 46  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 30  
LENGTH: 619  
TYPE: PRT  
ORGANISM: Actinoplanes sp.  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1)  
OTHER INFORMATION: V represents a non-standard initiator codon. It is expected that  
OTHER INFORMATION: at this position  
US-11-205-109-30

Query Match 5.5%; Score 109; DB 7; Length 619;  
Best Local Similarity 31.4%; Pred. No. 0.69;  
Matches 82; Conservative 22; Mismatches 93; Indels 64; Gaps 19;

QY 4 AIPGRIAGFVNNVMSVTLVMTFTWSLATMLPCTVG-----YPAVYVWLGGAG 56  
DB 85 SITDKLAGAFVPOQLSARV--FGFHQMSLA--LPQAVEGVIALVLYRAVRRM--HGRGA 139  
QY 57 VMLASLAHBLAHVAVRAGVSVE--SVTLMLFGVYALGGEAKTPKAFRIAPAPA 114  
DB 140 GLAAAGLPA---TPPIVSMGSHMEDGALTCLVLAADAFGA-AVTGSPARLALAGAW 195  
QY 115 TSLASLA-----TFGLAIT-LAG--VTPAIVISVAMWLA---TVNLLGL-FNLLP 160  
DB 196 IGLGQAMQAMVLPALVTVTYLAGAFVRAARAVVAAVAATLAVSLMLVATLILP 255  
QY 161 GAP--LDGRLVRAV--LMBRHGDSVRAGI---GAARAGRVVALVIALGLAEFVAGG- 211  
DB 266 GSHRWADGTTSGNAFAMVFGVNGFD--RAGIHVFGALTTG-----FTDGA 300  
QY 212 LVGVYWLAF-----IGWF 224  
DB 301 AAGGSWTALAADRLATQIGMW 321

RESULT 12  
US-10-858-730-121  
Sequence 121, Application US/10858730  
Publication No. US20050255568A1

GENERAL INFORMATION:  
; APPLICANT: Bailey, Richard B.  
; APPLICANT: Blomquist, Paul  
; APPLICANT: Doten, Reed  
; APPLICANT: Driggers, Edward M.  
; APPLICANT: Madden, Kevin T.  
; APPLICANT: O'Leary, Jessica  
; APPLICANT: O'Toole, George  
; APPLICANT: Trueheart, Joshua  
; APPLICANT: Walbridge, Michael J.  
; APPLICANT: Yorgey, Peter S.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID  
; TITLE OF INVENTION: PRODUCTION  
; FILE REFERENCE: 14184-030001  
; CURRENT APPLICATION NUMBER: US/10/858,730  
; CURRENT FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/475,000  
; PRIOR FILING DATE: 2003-05-30  
; PRIOR APPLICATION NUMBER: US 60/551,860  
; PRIOR FILING DATE: 2004-03-10  
; NUMBER OF SEQ ID NOS: 364  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 121  
; LENGTH: 334  
; TYPE: PRT  
; ORGANISM: Streptomyces coelicolor  
US-10-858-730-121

Query Match 5.4%; Score 106.5; DB 6; Length 334;  
Best Local Similarity 24.3%; Pred. No. 0.51;  
Matches 81; Conservative 32; Mismatches 111; Indels 109; Gaps 16;

Qy 2 RDA---IPGRI--AGFYVNVHVSYLVLMLFTMSLATMTPGVGYPVAVVWMLGAGA 56  
Db 61 RDAGQIAGLSQSGYLSVYATL-----GVSSGTALTI-----DGV 100  
Qy 57 VMLASLAEHLAAVAVR-----RAGVSVSVTLMLF----- 89  
Db 101 QPLVAGALGPIRLRGYVRGQMLGLMLGSAVTVDAGAAAGAVMMAVLVPFLGML 160  
Qy 90 --GGTALGGEAKTPKAFRIAF--GPAISLSTFGALATTAGVTPPAIVISVAMWL 146  
Db 161 SLVAATFLEGRTVRVAP--RVALLTHCATSAVL--FSGLAGLGAAPPA--GSSFWL 213  
Qy 147 ATVNLLGLFNLPGAPLDGRLVRAVLMRRHGS-----VRAGIGAPARARHV 195  
Db 214 ATAWLVV-----LPTFGYGLYMLIRSGITREVTMLMAPVAVMGALMFGEFP 265  
Qy 196 ALVIALGLAEFVAGLV--GGVWLAFIGWFIFAAREBEETRISTQULPAGVRVADAMT 252  
Db 266 G-VQPALGLAVGLAAVVVVRRGG-----GARRRPVRSAGDRPAAGGPTADQPT 313  
Qy 253 AQPHTAPGWINVEDFIQRYVLGERHSAYPVADR 285  
Db 314 NRPTDRPTAAGSTD-----RPTADR 333

RESULT 13  
US-11-087-099-6812  
; Sequence 6812, Application US/11087099  
; Publication No. US20060041961A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53450)B EP  
; CURRENT APPLICATION NUMBER: US/11/087,099  
; CURRENT FILING DATE: 2005-03-22  
; NUMBER OF SEQ ID NOS: 12464  
; SEQ ID NO 6812  
; LENGTH: 434  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis H37Rv  
US-11-087-099-6812

Query Match 5.3%; Score 105.5; DB 7; Length 434;  
Best Local Similarity 22.8%; Pred. No. 0.83;  
Matches 89; Conservative 39; Mismatches 137; Indels 125; Gaps 19;  
Qy 7 LGRAGFYVNVHVSVLVILMLFTMSLATMTPGVGYPVAVVWMLGAGAVMLASLLAH 66  
Db 66 LGGWAGF--SVGM---LYWTF--WVIYVGEBAVAG--KVLTYWTL----- 101  
Qy 67 ELAAVAVARAGVSVESVTLMLFGGYTALGGEAKTPRAAFRIAPGAPATSLATGEGAL 126  
Db 102 -----DAPLML-----ASLCMMMMTRATNLVSVSSGFEF 130  
Qy 127 AITLAGVTPPAIVISVAMMLATVNLLG---LENLPGAPLDGRLVRAVLMRRHGSVR 183  
Db 131 EFWFAGYKVAITV-----GELYVGTAFAGLLGHGMDFSNLT-----SAHGCFPP 175  
Qy 184 AGTGAARAGRVVALVIALGLAEFVAGLVGVGWLATIGWFIFAAREBEETRISTQULFA 243  
Db 176 DGVGAVFPAATVVA--IFSMGTGEVVT-----IAAARAPDQRAVQRAWS 217  
Qy 244 GV--RVADAMTAQHHTAP--GWINVEDFIQRYVLGERHSAYPVADRDSITGLVALROL 298  
Db 218 TVVARIYIFFGVSFLLTVILPMSLELGASPYVAALRHMGIGAGDQ---IMNAVULTAV 274  
Qy 299 RDVAPSRSTTSVGDIALPLHSVPTAPQEBPTALLERMAPLGRSPALVTESAVVGI- 357  
Db 275 LSCINSGLYTAS-----RMLFVLAARQEAR--AQLVKNRVGVPTPAIM--GSSVVGFL 324  
Qy 358 -----VTPSDV-----ARLIDYRL 372  
Db 325 CVIMAWVSPATVFEVLLNNSGAVLLEFVLL 354

RESULT 14  
US-11-087-099-3033  
; Sequence 3033, Application US/11087099  
; Publication No. US20060041961A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53450)B EP  
; CURRENT APPLICATION NUMBER: US/11/087,099  
; CURRENT FILING DATE: 2005-03-22  
; NUMBER OF SEQ ID NOS: 12464  
; SEQ ID NO 3033  
; LENGTH: 488  
; TYPE: PRT  
; ORGANISM: Rhodococcus erythropolis  
US-11-087-099-3033

Query Match 5.3%; Score 104.5; DB 7; Length 488;  
Best Local Similarity 24.3%; Pred. No. 1.1;  
Matches 56; Conservative 25; Mismatches 62; Indels 87; Gaps 12;

Qy 32 LATMLPSTGVGYP---AVYVWMLGAGAVMLASLAEHLAAVAVRAGVSVESVTL- 86  
Db 96 MSSALPAAAGGYTPARABMPW---GSPATGTALILEYSIAPAAITFGAVGESINLF 151  
Qy 87 -----WLFQGYTALGGEAKTPKAFRIAPAG--PATSLASTFGALATTAGVTPA 137  
Db 152 GITDGMWVYLAIVAI-----FTGHLTGAGELKXMFITITIALVGL----- 193  
Qy 138 IIVISVAMMLATVNLLGLFNLPGAPLDGRLVRAVLMRRHGSVRAIGIAGARGVVAL 197  
Db 194 VVFVVS-----AVGLF-----DSSULTD-----IAV 214  
Qy 198 VLIALGLAEFVAGLVGVGWLDA--FIGWFIFAA-----AREBEETRI 236  
Db 215 DTSAVGSSSFLPFGLL-GIMAAVFPFALIMFILAEGVPLAAEAREBEKEGV 263

RESULT 15

```

US-11-087099-9168
; Sequence 9168, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(513450) B EP
; CURRENT APPLICATION NUMBER: US/11/087, 099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 9168
; LENGTH: 557
; TYPE: prt
; ORGANISM: Thermoplasma acidophilum
US-11-087-099-9168

```

Query Match	5.3%	Score 104.5;	DB 7;	length 557;
Best Local Similarity	21.0%;	Pred. No. 1.3;		
Matches	61;	Mismatches	86;	Indels 103; Gaps 13,

```

QY 26 MLEFTSLMTMEGTGCGYPAVVYVYMLLACAGVALLSLAHLEIAHHV-----VARRAGVS 80
Db 39 MLEFAAGASMLMGPA-----AYLSMTI--GGILVLTVALYAEIAGMIIRGGAIVRGQYS 92
QY 81 VESVTLLMLFG-----GVTAIAGGEAKTPKPAFRIAPAGPATSLALATGGLAIIITLAC 132
Db 93 HGGLGFLPFGMAVYFLSASVPAIEAEGV-----ITYAGTYKGVHTPT--GVLT--TAVG 142
QY 133 VRTPAIVT-----SVAMW-----LATVNLILGIFENLLPGA 162
Db 143 ILLAAILMLGPPFLAYFGIKVMGKNTGNTGMWLLIIPSATVALLLVHFNIGNFPKLTGF 202
QY 163 PLDGG-----GRLYRAYILMR-----HGDSVRAGIGAAARVVALVT----- 199
Db 203 VTYGSSVFEAISGCVESYLCGRQALDYGGAKAPQRSVPATATLSVIGIALVALQ 262
QY 200 -IAGLAERVAGLYGVV-----LAIQGFITFAA 229
Db 263 VVFIQVNMMAAVGAPGQMSALSGGVTAAPATLASSAGIVFELSYFLPAD 313

```

Search completed: March 23, 2006, 06:29:39  
Job time : 11.431 secs



This Page Blank (usp.)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using SW model

Run on: March 23, 2006, 04:56:29 ; Search time 109.803 Seconds  
(without alignments)  
1652.628 Million cell updates/sec

Title: US-10-617-038-29

Perfect score: 2195  
Sequence: 1 MASSASDGTHERSAFRLSP...AALEQSGLDAPRTGRDRA 413

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_GeneSeq\_21: \*  
1: geneSeq1980s: \*  
2: geneSeq1990s: \*  
3: geneSeq2000s: \*  
4: geneSeq2001s: \*  
5: geneSeq2002s: \*  
6: geneSeq2003s: \*  
7: geneSeq2003bs: \*  
8: geneSeq2004s: \*  
9: geneSeq2005s: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2195	100.0	413	AD137308	Ad137308 M. tuberc
2	148	6.7	346	AAg20855	AAg20855 Arabidops
3	148	6.7	353	AAg20854	AAg20854 Arabidops
4	148	6.7	356	AAg20853	AAg20853 Arabidops
5	146.5	6.7	369	ADx77895	ADx77895 Plant ful
6	146.5	6.7	369	ADx68280	ADx68280 Plant ful
7	130	5.9	172	ABM92392	ABM92392 M. xanthu
8	126	5.7	582	ADCS2136	ADCS2136 Aeropyrum
9	121	5.5	338	ABM67065	ABM67065 Drosophyl
10	118.5	5.4	322	ADT59525	ADT59525 Plant pol
11	117.5	5.4	889	ABO60724	ABO60724 Klebsiell
12	112.5	5.1	409	ABO71490	ABO71490 Pseudomon
13	112	5.1	574	ABU34102	ABU34102 Protein e
14	109.5	5.0	1127	ABO78103	ABO78103 Pseudomon
15	106.5	4.9	691	ABG61612	ABG61612 Human DPR
16	106	4.8	452	ADY04294	ADY04294 Plant ful
17	105	4.8	343	AAU49953	AAU49953 Propionib
18	105	4.8	343	ABM46472	ABM46472 Propionib
19	104.5	4.8	742	ADG27327	ADG27327 Bacterial
20	104.5	4.8	742	ADG26598	ADG26598 Bacterial
21	104.5	4.8	742	ADG26593	ADG26593 Bacterial
22	102.5	4.7	4635	ADK56098	ADK56098 Streptomy
23	102	4.6	563	ABO68803	ABO68803 Pseudomon
24	101	4.6	518	ADT50839	ADT50839 Cancer re

25	100.5	4.6	586	9	AEb03469	AEb03469 Mycobacte
26	100.5	4.6	586	9	AEa79372	AEa79372 Novel M.
27	100	4.6	421	8	ADG95064	ADG95064 Human the
28	100	4.6	421	8	ADY67921	ADY67921 Biologica
29	99.5	4.5	424	8	ADY08484	ADY08484 Plant ful
30	99.5	4.5	981	9	ABM97613	ABM97613 M. xanthu
31	99	4.5	421	8	ADN61819	ADN61819 Human nov
32	98.5	4.5	748	9	ABM94678	ABM94678 M. xanthu
33	98.5	4.5	1178	9	ABM96130	ABM96130 M. xanthu
34	98	4.5	668	4	AAb79081	AAb79081 Corynebac
35	98	4.5	706	4	AAAG90125	AAAG90125 C. glutam
36	98	4.5	706	7	ADP13214	ADP13214 C. glutam
37	98	4.5	706	7	ADP13657	ADP13657 C. glutam
38	97	4.4	401	3	AAV77188	AAV77188 S. venezu
39	97	4.4	401	3	AAV80998	AAV80998 S. venezu
40	97	4.4	402	5	AAE24236	AAE24236 Streptomy
41	97	4.4	402	8	ADL91909	ADL91909 Streptomy
42	97	4.4	483	4	AAb93479	AAb93479 Human pro
43	97	4.4	3782	3	AAV77179	AAV77179 S. venezu
44	97	4.4	3782	5	AAE24228	AAE24228 Streptomy
45	97	4.4	3782	8	ADL91932	ADL91932 Streptomy

#### ALIGNMENTS

RESULT 1	AD137308	standard; protein; 413 AA.
ID	AD137308	
AC	AD137308;	
XX		
DR	22-APR-2004 (first entry)	
XX		
DE	M. tuberculosis low oxygen induced antigen Rv2727c SEQ ID NO:29.	
XX		
KW	mycobacterial infection; vaccine; tuberculosis;	
KW	Mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;	
XX	low oxygen induced antigen.	
OS	Mycobacterium tuberculosis.	
XX		
PN	WO2004006952-A2.	
XX		
PD	22-JAN-2004.	
XX		
PF	08-JUL-2003; 2003WO-DK000477.	
XX		
PR	13-JUL-2002; 2002DK-00001098.	
XX		
PA	(STAT-) STATENS SERUM INST.	
XX		
PI	Andersen P, Rosenkrands I, Stryhn A;	
XX	WPI; 2004-122778/12.	
DR	N-PSDB; AD137353.	
XX		
PT	Use of one or more polypeptides or their fragments, which are expressed	
PT	during the latent stage of the mycobacterial infection, and/or nucleic	
PT	acids encoding the polypeptides, for a therapeutic vaccine against	
XX	tuberculosis.	
PS	Claim 3; SEQ ID NO 29; 76pp; English.	
XX		
CC	The present invention describes polypeptides or their fragments, which	
CC	are expressed during the latent stage of a mycobacterial infection,	
CC	and/or nucleic acids encoding the polypeptides, which are useful for	
CC	creating a therapeutic vaccine against tuberculosis. Also described: (1)	
CC	a therapeutic vaccine against tuberculosis comprising one or more	
CC	polypeptides; (2) a method for treating an animal, including a human	
CC	being, with tuberculosis caused by virulent mycobacteria, e.g. by	
CC	Mycobacterium tuberculosis, M. africanum or M. bovis; (3) a method for	
CC	immunising an animal, including a human being, against tuberculosis	

CC caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis  
CC caused by virulent mycobacteria in an animal, including a human being;  
CC (5) a method for diagnosing previous or ongoing infection with a virulent  
CC mycobacterium; and (6) a method of diagnosing Mycobacterium tuberculosis  
CC infection in a subject. The polypeptides have antibacterial activities,  
CC and can be used in vaccines and in gene therapy. The polypeptides are  
CC useful for the manufacture of a therapeutic vaccine for treating an  
CC individual who is infected by a virulent mycobacterium, e.g. M.  
CC tuberculosis, and who is not vaccinated with BCG against tuberculosis.  
CC The present sequence represents a low oxygen induced antigen, which is  
CC used in the exemplification of the present invention.

XX Sequence 413 AA;

Query Match 100.0%; Score 2195; DB 9; Length 413;

Best Local Similarity 100.0%; Pred. No. 1.8e-220; Mismatches 0; Indels 0; Gaps 0;

Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASSASDGTHERSAFRLSPVLGAMGPPMHTGLVVAQSWRDYLGQPPDKLPAPRTAL 60

Db 1 MASSASDGTHERSAFRLSPVLGAMGPPMHTGLVVAQSWRDYLGQPPDKLPAPRTAL 60

QY 61 AAQAFRDIVILGKARPVSNHRYFERISQEVAAGLFYGRRWLEKPSGFFAQPPPLT 120

Db 61 AAQAFRDIVILGKARPVSNHRYFERISQEVAAGLFYGRRWLEKPSGFFAQPPPLT 120

QY 121 EVAARKVDRRRSFYRIFFDSGFTPHGEPGSGQRLSTYANNREVALLRHPEPPWLYC 180

Db 121 EVAARKVDRRRSFYRIFFDSGFTPHGEPGSGQRLSTYANNREVALLRHPEPPWLYC 180

QY 121 EVAARKVDRRRSFYRIFFDSGFTPHGEPGSGQRLSTYANNREVALLRHPEPPWLYC 180

Db 121 EVAARKVDRRRSFYRIFFDSGFTPHGEPGSGQRLSTYANNREVALLRHPEPPWLYC 180

QY 181 VHGTMGRAPIDLAVFRAMKLDDELGLNIWMPVLPMGPRGQGLPKGAVFGEEDYLDVH 240

Db 181 VHGTMGRAPIDLAVFRAMKLDDELGLNIWMPVLPMGPRGQGLPKGAVFGEEDYLDVH 240

QY 241 GTAQAVMIRRLSLMIRSOEESLIGLNGSLGCVITASLVASLEBGLACATLIGVPVADI 300

Db 241 GTAQAVMIRRLSLMIRSOEESLIGLNGSLGCVITASLVASLEBGLACATLIGVPVADI 300

QY 301 ELLGRHCGLRKHPRRHTVKNAPIGRMISPLSLTPLVPMGPRFYAGIADRLVHPRQV 360

Db 301 ELLGRHCGLRKHPRRHTVKNAPIGRMISPLSLTPLVPMGPRFYAGIADRLVHPRQV 360

QY 361 TRLMEHWKPEIWMYPGHTGFFQSGRPVRRFYQALLDSGLIDAPRTORDSA 413

Db 361 TRLMEHWKPEIWMYPGHTGFFQSGRPVRRFYQALLDSGLIDAPRTORDSA 413

RESULT 2  
AAG20855 ID AAG20855 standard; protein; 346 AA.

XX AAG20855; 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 23199.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.

XX Arabidopsis thaliana.

OS EPI033405-A2.

PN 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

XX 05-MAR-1999; 98US-0123180P.

XX 09-MAR-1999; 99US-0123548P.

XX 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138034P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140951P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.

PR	08-OCT-1999;	99US-0158233P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.

[illegible]

XX	Arabidopsis thaliana.	PR	29-JUN-1999;	99US-0140991P.
OS		PR	30-JUN-1999;	99US-0141287P.
XX	EPI033405-A2.	PR	01-JUL-1999;	99US-0141842P.
PN		PR	01-JUL-1999;	99US-0142154P.
XX		PR	02-JUL-1999;	99US-0142055P.
PD	06-SEP-2000.	PR	06-JUL-1999;	99US-0142390P.
XX		PR	08-JUL-1999;	99US-0142803P.
XX		PR	09-JUL-1999;	99US-0142920P.
PF	25-FEB-2000; 2000EP-00301439.	PR	12-JUL-1999;	99US-0142977P.
XX		PR	13-JUL-1999;	99US-0143542P.
XX		PR	14-JUL-1999;	99US-0143624P.
PR	25-FEB-1999;	PR	15-JUL-1999;	99US-0144005P.
PR	05-MAR-1999;	PR	16-JUL-1999;	99US-0144085P.
PR	09-MAR-1999;	PR	16-JUL-1999;	99US-0144086P.
PR	23-MAR-1999;	PR	16-JUL-1999;	99US-0144325P.
PR	25-MAR-1999;	PR	19-JUL-1999;	99US-0144331P.
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144332P.
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144333P.
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144334P.
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144335P.
PR	16-APR-1999;	PR	19-JUL-1999;	99US-0144335P.
PR	19-APR-1999;	PR	20-JUL-1999;	99US-0144335P.
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144335P.
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144335P.
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144335P.
PR	28-APR-1999;	PR	21-JUL-1999;	99US-0144814P.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145086P.
PR	04-MAY-1999;	PR	21-JUL-1999;	99US-0145086P.
PR	05-MAY-1999;	PR	22-JUL-1999;	99US-0145087P.
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145089P.
PR	07-MAY-1999;	PR	22-JUL-1999;	99US-0145182P.
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145185P.
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145224P.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145276P.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145913P.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145918P.
PR	18-MAY-1999;	PR	27-JUL-1999;	99US-0145919P.
PR	19-MAY-1999;	PR	28-JUL-1999;	99US-0145951P.
PR	20-MAY-1999;	PR	02-AUG-1999;	99US-0146386P.
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146388P.
PR	24-MAY-1999;	PR	02-AUG-1999;	99US-0146389P.
PR	25-MAY-1999;	PR	03-AUG-1999;	99US-0147038P.
PR	27-MAY-1999;	PR	04-AUG-1999;	99US-0147204P.
PR	28-MAY-1999;	PR	04-AUG-1999;	99US-0147302P.
PR	01-JUN-1999;	PR	05-AUG-1999;	99US-0147260P.
PR	03-JUN-1999;	PR	05-AUG-1999;	99US-0147303P.
PR	04-JUN-1999;	PR	06-AUG-1999;	99US-0147303P.
PR	07-JUN-1999;	PR	06-AUG-1999;	99US-0147416P.
PR	08-JUN-1999;	PR	09-AUG-1999;	99US-0147435P.
PR	10-JUN-1999;	PR	09-AUG-1999;	99US-0147935P.
PR	10-JUN-1999;	PR	10-AUG-1999;	99US-0148171P.
PR	14-JUN-1999;	PR	11-AUG-1999;	99US-0148319P.
PR	16-JUN-1999;	PR	12-AUG-1999;	99US-0148341P.
PR	16-JUN-1999;	PR	13-AUG-1999;	99US-0148565P.
PR	17-JUN-1999;	PR	13-AUG-1999;	99US-0148684P.
PR	18-JUN-1999;	PR	16-AUG-1999;	99US-0149368P.
PR	18-JUN-1999;	PR	17-AUG-1999;	99US-0149175P.
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149442P.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149722P.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723P.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149929P.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149902P.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149930P.
PR	18-JUN-1999;	PR	25-AUG-1999;	99US-0150566P.
PR	18-JUN-1999;	PR	26-AUG-1999;	99US-0150884P.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151065P.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151066P.
PR	21-JUN-1999;	PR	27-AUG-1999;	99US-0151080P.
PR	22-JUN-1999;	PR	30-AUG-1999;	99US-0151303P.
PR	23-JUN-1999;	PR	31-AUG-1999;	99US-0151438P.
PR	23-JUN-1999;	PR	01-SEP-1999;	99US-0151930P.
PR	24-JUN-1999;	PR	07-SEP-1999;	99US-0152363P.
PR	28-JUN-1999;	PR	10-SEP-1999;	99US-0153070P.

PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154033P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155133P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 6.7%; Score 148; DB 3; Length 353;  
Beet Local Similarity 27.7%; Pred. No. 3,1e-06;

Matches 93; Conservative 38; Mismatches 133; Indels 72; Gaps 20;

QY 82 NHRVERISQEVAGLEFYGNRRMLEKPSGFAPPPLEVAVRKYKDRRSFYR--TFP 139  
DB 39 NLELLERAVQRLFP-LEVOG-QNM-----PPLVPRPVKRTVMTKTATLREGVF- 85  
QY 140 DSGFTPHDGE-----PGSQRMVSTYANNREYALLLRHPEPRMVLVCVHTEMGRAPLDLA 194  
DB 86 ---GTPCADELTAALPPESR-----TA--RVAMLVKQVPPQGMACVHLAGDHTYD-- 134  
QY 195 VFRAKMKHDEL-GLNTVMPVL--PMAGPRGGLPFGA-VFPGEDVLDDVHGTAAQAVMIR 250  
DB 135 --RRRLRGGLVKNQIATMVLSPFYGQRPRPLQCGARLLCVSDLL--LLGGA--TIESR 189  
QY 251 RLWSIRSQEBSLTLGLNGLSIGYIASIVASLEBGLACALIGVPADLIBLGRH---- 306  
DB 190 SLIHMLDTBEGKMGVCGLSGVHASMVCSLHP-----TVATL-PLLSHSAV 240  
QY 307 --C-----GLRHK-DPRRHTVMAEBIGRMISPLSLT-----PLVPMGRFTYA 347  
DB 241 AFGCEILKYGTAWMALREBLAAQKITMTLDEVREBRMNVLSITDTRPPIPKNPDAVLFV 300  
QY 348 GIADRLVHPRBOYTLMEHMGKPELIVYPGGHTGTF 383  
DB 301 AATDGYIPKHSVLELQKAMPGEVRAVWTGGHVSSF 336

RESULT 4  
AAG20853  
ID AAG20853 standard; protein; 356 AA.  
XX  
AC AAG20853;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 23197.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hydridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123160P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 01-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137750P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.



QY 251 RLSTIRSOEBSLIGLNGLSIGYIASLVASLEBGLACAIIGVADLIIEILGRH---- 306  
DB 193 SLIHLMDTEBGGKMGVCGLSMGVNASVWGSILHP-----TVAATL-PELSPHSNAV 243  
QY 307 ---C-----GLRHK-DPRRHTVMAEPIGSMISPLSLT-----PLVMPGRFIYA 347  
DB 244 AFCEGILKKYGTWELRLELAQAQITMTLDEVERRRNVSLDVTTRFPFIPKONPAVILFV 303  
QY 348 GIADRLVHPREQVTRLMWHMGKPEIWMYFGHTGFF 383  
DB 304 AATDGYIPKHSVLELQKAMPSESVRWVTGHSVSF 339  
RESULT 5  
ADK77895  
ID ADK77895 standard; protein; 369 AA.  
AC ADK77895;  
DT 21-APR-2005 (first entry)  
DE Plant full length insert polypeptide seqid 47261.  
XX plant protectant; plant growth regulant; gene therapy; plant;  
XX recombinant DNA construct; physical array; plant breeding marker;  
XX cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
XX extreme osmotic condition; pathogen tolerance; pest tolerance;  
XX growth rate; cell cycle pathway; disease resistance;  
XX galactomannan production; lignin production; plant growth regulator;  
XX yield; plant growth; plant development; seed oil; protein yield;  
XX protein content.  
XX Unidentified.  
OS US2004034888-A1.  
PN 19-FEB-2004.  
PD 28-APR-2003; 2003US-00425114.  
PP 06-MAY-1999; 99US-00304517.  
PR 05-NOV-2001; 2001US-00985678.  
XX  
PA (LIUJ/) LIU J.  
PA (ZHOU/) ZHOU Y.  
PA (KOVA/) KOVALIC D K.  
PA (SCRE/) SCREEN S E.  
PA (TABAS/) TABASKA J E.  
PA (CAOY/) CAO Y.  
XX  
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JB, Cao Y;  
XX WPI; 2004-180133/17.  
DR  
XX New recombinant DNA construct, useful for improving plant tolerance to  
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
PT pests, for conferring increased resistance to plant disease, or for  
XX improving yield.  
XX  
PS Claim 1; SEQ ID NO 47261; 15pp; English.  
XX  
XX The invention describes a recombinant DNA construct comprising a  
CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US patent office at  
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring  
CC increased resistance to plant disease, for producing galactomannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of

CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This is the amino acid sequence of a plant full length insert  
CC polypeptide that can be used in the recombinant DNA construct of the  
CC invention.  
XX  
SQ Sequence 369 AA;  
Query Match 6.7%; Score 146.5; DB 8; Length 369;  
Best Local Similarity 24.6%; Pred. No. 4.8e-06;  
Matches 59; Conservative 31; Mismatches 81; Indels 69; Gaps 8;  
QY 208 NIWPEVL--PMHGRGGLPRGA-----VFGEDVLDVHGTAQVMDIRILSIR 257  
DB 149 NIATWLESFYGGGRPSMGQAKLCQVSDLLIGKATIDBA-----RSLIWMIQ 198  
QY 258 SOEBSLIGLNGLSIGYIASLVASLE-----RGL----- 287  
DB 199 NEAGYKMGICGLSMGVHAAVMSLHPTVATLPELAPHSAAVVPFCGVYKATAMDAL 258  
QY 288 --ACAILGVPAVDLIIEILGRHCGLRHKDPRRHTVMAEPIGSMISPLSLT-----PLVPM 340  
DB 259 RKDAAVLTIQDVTLLAEDAAQSGITIEQV-----DRLSVSLTDVTRPVPKX 308  
QY 341 PGRFIYAGIADRLVHPREQVTRLMWHMGKPEIWMYFGHTG--FFQSRPVRFPQALAEQ 398  
DB 309 FOAVIFVATDGYITPRHSVWELQKAMPSESVRWVTGHSVSFFLHNDAPRAKAIYDALDR 368  
RESULT 6  
ADK68280  
ID ADK68280 standard; protein; 369 AA.  
AC ADK68280;  
DT 21-APR-2005 (first entry)  
DE Plant full length insert polypeptide seqid 39123.  
XX  
XX plant protectant; plant growth regulant; gene therapy; plant;  
XX recombinant DNA construct; physical array; plant breeding marker;  
XX cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
XX extreme osmotic condition; pathogen tolerance; pest tolerance;  
XX growth rate; cell cycle pathway; disease resistance;  
XX galactomannan production; lignin production; plant growth regulator;  
XX yield; plant growth; plant development; seed oil; protein yield;  
XX protein content.  
XX Unidentified.  
OS US2004034888-A1.  
PN 19-FEB-2004.  
PD 28-APR-2003; 2003US-00425114.  
PP 06-MAY-1999; 99US-00304517.  
PR 05-NOV-2001; 2001US-00985678.  
XX  
PA (LIUJ/) LIU J.  
PA (ZHOU/) ZHOU Y.  
PA (KOVA/) KOVALIC D K.  
PA (SCRE/) SCREEN S E.  
PA (TABAS/) TABASKA J E.  
PA (CAOY/) CAO Y.  
XX  
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JB, Cao Y;  
XX WPI; 2004-180133/17.  
DR  
XX New recombinant DNA construct, useful for improving plant tolerance to  
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or



PT pests, for conferring increased resistance to plant disease, or for  
PT improving yield.  
XX  
PS Claim 1; SEQ ID NO 39123; 15pp; English.  
XX  
CC The invention describes a recombinant DNA construct comprising a  
CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US patent office at  
CC ftp://seqdata.uspto.gov/sequence.html?docID:2004034888. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring  
CC increased resistance to plant disease, for producing galactomannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This is the amino acid sequence of a plant full length insert  
CC polypeptide that can be used in the recombinant DNA construct of the  
CC invention.  
XX  
SQ Sequence 369 AA;  
XX  
Query Match 6.7%; Score 146.5; DB 8; Length 369;  
Best Local Similarity 24.6%; Pred. No. 4,8e-06;  
Matches 59; Conservative 31; Mismatches 81; Indels 69; Gaps 8;  
OY 208 NIVPEVL--PMHPRGGLPKGA-----VFGEVDVHGTAAQVMDIRLLSIR 257  
DB 149 NIATVLESPPYGGRRPSMGRGAKQCVSDLLLGKATIDEA-----RSILYWLQ 198  
OY 258 SOBEESLGLNGSLGCVIASLVASLE-----BGL----- 287  
DB 199 NEAGYGRMGICGLSGVYAAVWSLHPTPVATLPLFAPHSANVPCEGVKYATAMDAL 258  
OY 288 --ACAILGVPAVDLIELGRHGRKDPRRHTVMAEPGIMISPLT-----PLVPM 340  
DB 259 RKDAAVLTQDVTILAEADAQSKGITIEQV-----DLRQVSLTLDTVTRFPVPM 308  
OY 341 PGRFTYAGIADRLVHPRQVTRFLWEHWGKPEIIVWYFGGHTG--FQSPRPVRREFVALLEQ 398  
DB 309 FQAVIFVGATDDGYIPRHSVWELQKAMPGESEVRWVTGGVSSFFLHNDAPFRKAIYDALDR 368  
RESULT 7  
ABM92392  
ID ABM92392 standard; protein; 172 AA.  
XX  
AC ABM92392;  
XX  
DT 02-JUN-2005 (first entry)  
XX  
DE M. xanthus protein sequence, seq id 11591.  
XX  
KM Transgenic plant; DNA replication; gene regulation; gene expression.  
XX  
OS Myxococcus xanthus.  
XX  
PN US6833447-B1.  
XX  
PD 21-DEC-2004.  
XX  
PF 10-JUL-2001; 2001US-00902540.  
XX  
PR 10-JUL-2000; 2000US-0217883P.  
XX  
PA (MONS ) MONSANTO TECHNOLOGY LLC.  
XX  
FI Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;  
XX

DR WPI; 2005-028716/03.  
XX  
XX New substantially purified Myxococcus xanthus nucleic acid molecule  
PT encoding a nitrite reductase, useful for determining gene expression,  
PT identifying mutations in a gene of interest, and for constructing  
PT mutations in a gene of interest.  
XX  
PS Example 2; SEQ ID NO 11591; 25pp; English.  
XX  
CC The invention relates to a substantially purified nucleic acid molecule  
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a  
CC recombinant DNA construct for expression of a nitrite reductase gene in a  
CC plant cell, and a plant cell comprising the recombinant DNA construct.  
CC The nucleic acid is useful for determining gene expression, identifying  
CC mutations in a gene of interest, and for constructing mutations in a gene  
CC of interest. Sequences given in records for SEQ IDs 9692-16825 represent  
CC a group of 7134 Myxococcus xanthus proteins and peptides. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from USPTO  
XX  
SQ Sequence 172 AA;  
XX  
Query Match 5.9%; Score 130; DB 9; Length 172;  
Best Local Similarity 30.4%; Pred. No. 8.7e-05;  
Matches 45; Conservative 21; Mismatches 72; Indels 10; Gaps 7;  
OY 248 DIRRLSWIRSOBEESLGLNGSLGCVIASLVAS--LEEGIACAILGV--PVADLIE-L 302  
DB 8 EARLLAMLRN--EGHARLVGAGVSMGGYMAALTAAVPEPLAVALAAGASFPVFNQGL 66  
OY 303 LGRHGLRHKD--PRRHVYKMAEPYGRMISPLTFLVP--MPGFTYAGIADRLVHPRQ 359  
DB 67 LWSIAFALLDGPFRDBEQARSRLGRIFDLANLARFPFPQPEAAVAVACRRDGFVPDE 126  
OY 360 VTRLEHWGKPEIIVWYFGGHTG--TGFGQSR 386  
DB 127 TLALHAMPRSELKRWVDAGHVTLFTER 154  
RESULT 8  
ADC52136  
ID ADC52136 standard; protein; 582 AA.  
XX  
AC ADC52136;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Aeropyrum pernix hyperthermophilic esterase, SEQ ID 2.  
XX  
KM Enzyme; hyperthermophilic esterase; peptidase; esterase; ester.  
XX  
OS Aeropyrum pernix.  
XX  
PN JP2002253244-A.  
XX  
PD 10-SEP-2002.  
XX  
PF 28-FEB-2001; 2001JP-00054949.  
XX  
PR 28-FEB-2001; 2001JP-00054949.  
XX  
PA (DOKU-) DOKURITSU GYOSEI HOJIN SANGYO GIJUTSU SO.  
XX  
PN WPI; 2003-460507/44.  
XX  
DR WPI; 2003-460507/44.  
XX  
PT Hyperthermophilic esterase, a polynucleotide, a polypeptide, a  
PT recombinant vector, a transformant, preparation of the hyperthermophilic  
PT esterase, use of the hyperthermophilic esterase.  
XX  
PS Claim 4; SEQ ID NO 2; 10pp; Japanese.  
XX  
CC The present sequence is a hyperthermophilic esterase (I) from Aeropyrum  
CC pernix, which has peptidase and esterase activity at a temperature range



XX PS Claim 2; SEQ ID NO 9602; 14pp; English.  
XX  
CC The invention relates a recombinant DNA construct comprising a  
CC polynucleotide having any of 5544 nucleotide sequences (cdnas SEQ ID NO:  
CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,  
CC Arabidopsis, wheat and rape but the specification does not indicate which  
CC sequences is derived from which organism. Also included is a method of  
CC producing a plant having an improved property, comprising transforming a  
CC plant with a recombinant DNA construct comprising a promoter region  
CC functional in a plant cell operably joined to a polynucleotide encoding a  
CC polypeptide associated with the property, and growing the transformed  
CC plant. The property is selected from improving plant cold tolerance, for  
CC manipulating growth rate in plant cells by modification of the cell cycle  
CC pathway, for improving plant drought tolerance, for providing increased  
CC resistance to plant disease, for galactomanan production, for production  
CC of plant growth regulators, for improving plant heat tolerance, for  
CC improving plant tolerance to herbicides, for increasing the rate of  
CC homologous recombination in plants, for lignin production, for improving  
CC plant tolerance to extreme osmotic conditions, for improving plant  
CC tolerance to pathogens or pests, for yield improvement by modification of  
CC photosynthesis, for modifying seed oil yield and/or content, for  
CC modifying seed protein yield and/or content, for yield improvement by  
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
CC and for yield improvement by providing improved plant growth and  
CC development under at least one stress condition. The polynucleotide may  
CC also encode a plant transcription factor. The methods and compositions of  
CC the present invention are useful in the field of biotechnology and  
CC genetics, in particular for producing transgenic plants with improved  
CC biological characteristics such as increased yield, improved nitrogen  
CC flow, increasing plant tolerance to cold or heat, improving plant  
CC tolerance to extreme osmotic and drought conditions, and improving plant  
CC tolerance to plant pests or pathogens. They can also be used in physical  
CC arrays of molecules, plant breeding markers, computer-based storage and  
CC analysis systems. The present sequence is one of the 5544 plant protein  
CC sequences of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?docID=20040216190.  
XX  
SQ Sequence 322 AA;  
Query Match 5.4%; Score 118.5; DB 8; Length 322;  
Best Local Similarity 25.3%; Pred. No. 0.0034;  
Matches 57; Conservative 25; Mismatches 76; Indels 67; Gaps 10;  
QY 208 NINPVLV--PMHGRGGLPRKGA-----VPGEDVLDDVHGTAGVMDIRLLSGIR 257  
DB 130 NIATVLESPPYGGRRBPMQHGAKLQCVSDLLLGKATIDDA-----RSLLYWLO 179  
QY 258 SQEESLVLGLNGLSGYIASLVLEGLACAILGV-----VADLIELGRHC--- 307  
DB 180 AEPYGMKGICGLSMGCVHAAAGSL---LSTPIPLPLPHSAVVPFCGLVRHATAW 236  
QY 308 -GLRH-----KDRPRHTVKMA-----EPIGRMISPLSLPLVPMRGRTYAGIADRLVH 355  
DB 237 EALREDDAALAKDASTLEDAAAGITTEQVXROIT-----DDGYI 276  
QY 356 PREQVTLMEHWKGEPIVWPGCHTGF--QSRPVRFPVQALPQ 398  
DB 277 PRHSVMEIQKAWPGESEVRAWVTGHHVSSFLHNDSPKRAIVDALDR 321  
RESULT 11  
AB060724  
ID AB060724 standard; protein; 889 AA.  
XX  
XX AC AB060724;  
XX  
XX DT 29-JUL-2004 (first entry)  
XX  
DB Klebsiella pneumoniae polypeptide seqid 7241.

XX  
KM Recombinant expression vector; transcription regulatory element;  
XX Klebsiella pneumoniae protein; antibacterial; Vaccine.  
XX  
OS Klebsiella pneumoniae.  
XX  
PN US6610836-B1.  
XX  
PD 26-AUG-2003.  
XX  
PF 27-JAN-2000; 2000US-00489039.  
XX  
PR 29-JAN-1999; 99US-0117747P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
PI Breton GL, Osborne M;  
DR WPI; 2003-895346/82.  
DR N-PSDB; ACH94275.  
PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
PT preparing a vaccine composition against Klebsiella pneumoniae.  
PS Disclosure; SEQ ID NO 7241; 932pp; English.  
XX  
XX The invention describes a new isolated nucleic acid encoding a Klebsiella  
XX pneumoniae polypeptide. Also described are: a recombinant expression  
XX vector comprising the nucleic acid, operably linked to a transcription  
XX regulatory element; and a cell comprising the recombinant expression  
XX vector. The nucleic acid is useful for preparing a vaccine composition  
XX against Klebsiella pneumoniae. This is the amino acid sequence of a  
XX Klebsiella pneumoniae polypeptide of the invention  
XX  
SQ Sequence 889 AA;  
Query Match 5.4%; Score 117.5; DB 7; Length 889;  
Best Local Similarity 22.8%; Pred. No. 0.018;  
Matches 100; Conservative 59; Mismatches 158; Indels 121; Gaps 26;  
QY 20 PVLGAMG-----PRMHTGLVVAQSWRD--YLGGQDPDKLPIARPTTLAAQAFDEIVLL 72  
DB 240 PVITPALNNAAGFIKPVTRQGL-----SPADVFPDPDEPKR-----REASRPFDQVGP 289  
QY 73 GL-----KARPVSNHNVFERISQEVAAGLEFYGNRRMLEKPSGFFAQPPELVAV 124  
DB 290 GLVYLEAQMGMGKTEALVYAQCWLVOEKATGI-----YFALPTOLTS--- 332  
QY 125 RKVYDRRRSFYRTFFDSGFTPHQDEPSQRMWLYTANNREYALLRHPRRPMVCHGT 184  
DB 333 NKTYDRNAFLHQIV-STETP-----CHSLLL--HSGAWLM--DT 367  
QY 185 EMGRAPLDLVFRAWKHDELGL-----NIVMPVLPM-HG-PRGGGLPRGAVFPG 232  
DB 368 EMGE---EGSPGAGFPHRRKRGILLAPAVGTIOALMAVNVHGFRAAGLAKV--- 421  
QY 233 EDVLDDVH-----GT-AQAVMDIR-----LLSWIRSQ-BEESLGLNGSLGYIA 277  
DB 422 --ILDEVHTYDLYGTITNALVFLRQIDCTVILISATLSQTRDALLQOSTSEAVPLI 479  
QY 278 SLVASLEEGLACAILGVPAVDLELGRHCGLNHKORRHTVMAEPDGMISPLST-- 335  
DB 480 TAAPSABERERGLVEIGVPTVENTVTYIHSQ--RKDEPARREALRRRELGGQVLTIENTIA 537  
QY 336 ----PLVPMRGRTYAGIADRLVHPR---EQVTLMEHWKGEPIVWPGCHTGFQSRPV 388  
DB 538 EAQOTYIDLASRAVEACETGTLHSRTPQHRNRHEGRW-----VALY--GRAHWPKQK 591  
QY 389 RRFV--QAALEQSLDA 404  
DB 592 GRILVGTQVLEQSLDIDA 609

```
RESULT 12
AB071490
ID ABO71490 standard; protein; 409 AA.
XX
AC ABO71490;
XX
DT 29-JUL-2004 (first entry)
XX
DE Pseudomonas aeruginosa polypeptide #3665.
XX
KM Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.
XX
OS Pseudomonas aeruginosa.
XX
PN US6551795-B1.
PD 22-APR-2003.
XX
PF 18-FEB-1999; 99US-00252991.
XX
PR 18-FEB-1998; 98US-0074788P.
PR 27-JUL-1998; 98US-0094190P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX
DR WPI; 2003-615309/58.
XX
DR N-PSDB; ABD05061.
XX
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure: SEQ ID NO 20236; 455bp; English.
XX
CC The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABO67826-
CC ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 409 AA;
Query Match 5.1%; Score 112.5; DB 7; Length 409;
Best Local Similarity 24.7%; Pred. No. 0.02;
Matches 94; Conservative 33; Mismatches 111; Indels 143; Gaps 23;
QY 105 WLEKPSGFFAOPRPTEVAVRKVDKRRSPYRIEFDG---FTRHPGESSQRMWLY-- 158
DB 66 WDEYATGVRIRHNPAT-----ADNRR---RLFALPEPRATLLEGHPG-GGSVPWGEYAT 114
QY 159 -----TANNEVALLRHRP-----RPWLVCV----- 181
DB 115 GVIRRYRATADNRGRFLADNVRCKRATGMPRCMTATGATRCSPAVALLRHNRAOV 174
QY 182 -----HGTMRGRARLDLAVFRAMKLDDELINIVM--PVL-----PMHGPR-----GQ 222
DB 175 DTDNHRVVUG--GVRILGVGVAVLLRGRGLDGVVONHPRVLGVNDVRLNHAAYLARAGVGE 233
QY 223 GLRKGAVERGSDV-----LDDVH-----GTAAQAVDIR--RLLSWIR 257
```

```
DB 234 GVVAVAGLPSRHVIGIALSCARLTRHOGVAGADVDHVRVFIGIGIEVADDOAIRLLGARR 293
QY 258 SQEBSLSLGLNGSLGGYIASVLSLEGLACAILGVPAVDLILGRHGLR--HKDPR 315
DB 294 IARQVHQGLGGARPG-----QVAVVALAVAGVRVA---VAGBALRLQVHRDQG 339
QY 316 RH-TVKRAEPIGRMISPLSTPLVPMGPRFPIYAGIADRLVHPRQVTRLMH---WGKP 370
DB 340 AFAAIDLDLEGLGQ-----GRTV-AGVVELRVH-----RTVEHRGSGIGAD 378
QY 371 EIVWYPGGHTGFF--QSRPVR 389
DB 379 LVAAIDPFGNFVAQAQLRPVR 399
RESULT 13
ABU34102
ID ABU34102 standard; protein; 574 AA.
XX
AC ABU34102;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #19629.
XX
KM Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Mycobacterium avium.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US09107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyckind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
XX
PS N-PSDB; ACA37972.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
Claim 25; SEQ ID NO 62026; 1766bp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 623 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
```

CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 574 AA;

Query Match 5.1%; Score 112; DB 6; Length 574;  
Best Local Similarity 22.5%; Pred. No. 0.036; Mismatches 148; Indels 168; Gaps 24;  
Matches 105; Conservative 46; Mismatches 148; Indels 168; Gaps 24;

QY 23 SGAMGPFMTGLY-----VAQSMRDYLGQOPDKLPAPRTALAAQFNDIYLGLKA 76  
DB 13 AGPLRPFADAGIFEAADIHVAQRLTALTGESDDRVALA---VALLVRLRGSGVCVDLRA 69  
QY 77 RRPVSNRVERISQEV-----AAG-----LEFYG-----NRRL 106  
DB 70 -----VPAQGAADLPMPAAGDWLAAYASPLGAPPVLRFFGDLVYFDRWL 117  
QY 107 EK-----PSGFFAQ-----PPPLTEVAVRKVDORRSFYRIFPDSGFTPH 147  
DB 118 EEEVCTDLALASAPAGVSSCYERLFPFGYER-----QRAARIVSQALTYLT 168  
QY 148 GEPSSQRLWLTANNREYALL---RHPEPRMLVCVHTGEMGA----- 189  
DB 169 GGPETGK---TTTVARLLALLVEQERAGEBRPXIALAPGKAARLAEAIAEIEHLD 225  
QY 190 PLD---LAVPRAMLDHDELGINIYMPVPMHGPRGQGLPKKAVRGEDVDDNGTAQAV 246  
DB 226 PADARLAGLTGTLRHRLGPRPDTSVAFKHN-RGNRLPHDI---VDETSMVSLTM 279  
QY 247 WDIRRLSWIRSGEESLIGNGLSLGGYIASLVASLEGLACAILGVPAVDLIETLGRH 306  
DB 280 --MARLAEAVRPDRRLIVG-----DDQLASVAGAX-----LADLVGGLAGR 321  
QY 307 CGLRHKDPRRHTVKAER--IGRMISPLSTPLVPMGRFTYAGIADRLVHPREQVTRLM 364  
DB 322 AGVR-----VALATPHRFGSAIGALAAA-----IRAXDADRIV----- 355  
QY 365 EHMGRPELVMPGHTGTFQGS-RPVRRVQA-----ALESGILDA 404  
DB 356 -----ELTLAGGEHIEWDSEPRADRLREVLVSHALTRSAALLGA 396

RESULT 14

ID ABO78103 standard; protein; 1127 AA.

XX ABO78103;

DT 29-JUL-2004 (first entry)

DE Pseudomonas aeruginosa polypeptide #10278.

KM Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.

XX Pseudomonas aeruginosa.

XX US6551795-B1.

XX 22-APR-2003.

PF 18-FEB-1999; 99US-00252991.

XX

PR 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
PA (GENO-) GENOME THERAPEUTICS CORP.  
PI Rubentfield MJ, Nolling J, Deloughery C, Bush D;  
DR WPI; 2003-615309/58.  
DR N-PSDB; ABD11674.

PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.

XX Disclosure; SEQ ID NO 26849; 455bp; English.

XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnosis and  
XX therapy of pathological conditions, as molecular targets for diagnosis,  
XX prophylaxis and treatment of pathological conditions resulting from a  
XX bacterial infection, for evaluating a compound, such as a polypeptide,  
XX for the ability to bind a *P. aeruginosa* nucleic acid, as components of  
XX effective antibacterial targets, as targets for antibacterial drugs,  
XX including anti-*P. aeruginosa* drugs, as templates for recombinant  
XX production of *P. aeruginosa*-derived peptides or polypeptides, as target  
XX components for diagnosis and/or treatment of *P. aeruginosa*-caused  
XX infection, and in detection of *P. aeruginosa* sequences or other sequences  
XX of Pseudomonas species using biotech technology. Sequences ABO67826-  
XX ABO84396 represent *P. aeruginosa* polypeptides of the invention. Note: The  
XX sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format from USPTO at  
XX seqdata.uspto.gov/sequence.html

SQ Sequence 1127 AA;

Query Match 5.0%; Score 109.5; DB 7; Length 1127;  
Best Local Similarity 23.5%; Pred. No. 0.17; Mismatches 100; Indels 137; Gaps 18;  
Matches 84; Conservative 36; Mismatches 100; Indels 137; Gaps 18;

QY 123 AVRKVDORRRSFYRIFPDSGFTPHGPGSQRLWSTYANNREYALLRHPEPR---MLV 179  
DB 27 AARSIPERR-----PGRPCNR-----CTWTRSVSPLPDGSAPFASAWM-- 65  
QY 180 CNGHTEMGAPDLAVFRANKH-DEGLNITVMPVPMHGPRGQGLPKGAV-----FPG 232  
DB 66 ----SRSGRCSCPYSCTSRLEFVDDERLELRIGIPVAHEGGQIGQEVVAHAGRAPV 121  
QY 233 EDVLDDVHGTAAQVMDIRLLSWIRSGEESTL---IGNGLSLGGYIASLVASLEGLA 288  
DB 122 DRDADVHGLA-----VHLQRADALGHHRRRLDVAVGGDL-DLVA---GID 164  
QY 289 CALIGVPVADLIETLGRHCGLRHKDPRRHTVKAERPGRMISPLSTPLV-----PMGR 343  
DB 165 AQLIGRLADLDDELG-----LGDRIQPAVLGVPEVGEPRVSGR 204  
QY 344 FI-----YAGIADRL-----VHPRECVTL-----WMHMKPRET 372  
DB 205 RVGELVGRAEQFVULLENPCRRYADRLAVVAVOGVHPRDRLERLVULGERSAFGLVDGEQ 264  
QY 373 VMYRPG-----HTGPGQSRPVRRFVQAL-----EOSGLDAPRTQRP 410  
DB 265 ASHPFGVNDERVHGRF-----RRYVGLVGVGAAPGLAVPRPHQALLKVPRLAD 315

RESULT 15

ID ABG61612 standard; protein; 691 AA.

XX ABG61612;

DT 12-AUG-2002 (first entry)

DB Human DRRP-3 splice variant #2.

XX	Human; serine protease; dipeptidyl peptidase IV-related protein; DPP; DPP-IV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain; diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke; heart failure; hypertension; urinary retention; osteoporosis; cancer; ulcer; allergy; cancer; psychotic disorder; neurological disorder; dyskinesia; reproductive disorder; inflammatory disorder; metabolic disorder.
XX	
OS	Homo sapiens.
XX	
FN	WO200231134-A2.
XX	
PD	18-APR-2002.
XX	
PF	12-OCT-2001; 2001WO-US031874.
XX	
PR	12-OCT-2000; 2000US-0240117P.
XX	
PA	(FERR ) FERRING BV.
XX	
P1	Oi S, Akinsanya KO, Riviere PJ, Junien J;
XX	
DR	WPI; 2002-444178/47.
XX	
DR	N-PSDB; ABR83343.
XX	
PT	New dipeptidyl peptidase IV-related proteins and nucleic acids encoding the proteins, useful for treating e.g. fungal, bacterial, protozoan and viral infections, cancers, allergies, neurological disorders, or pain.
XX	
PS	Disclosure; Page 107-109; 113pp; English.
XX	
CC	The present invention relates to the isolation of novel human serine proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP) and nucleic acids encoding them are useful for treating infections such as fungal, bacterial, protozoan and viral infections, particularly infections caused by human immunodeficiency virus (HIV-1 or HIV-2), pain, diabetes, precocious puberty, infertility, obesity, anorexia, bulimia, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, angina pectoris, stroke, ulcers, asthma, allergies, cancers, migraine, vomiting, psychotic and neurological disorders (e.g. anxiety, dementia, or schizophrenia), and dyskinesias. These may also be used in discovering therapeutic agents for the treatment of reproductive, inflammatory and metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins
CC	
XX	
SQ	Sequence 691 AA;
XX	
Query Match	4.9%; Score 106.5; DB 5; Length 691;
Best Local Similarity	18.2%; Pred. No. 0.18;
Matches	63; Conservative 63; Mismatches 106; Indels 115; Gaps 14
9	THRSAPFLSPVLSGAMGPFMAHTGYYVAQSRDYLGQDPDKLTARPTLALAAQAFDE 68
288	SRDSGKFMVTVVKQSGRGGEFHTIAMFLIQSSEGI-----TVRLHTSGNNEV 335
69	IVLGLKARRPVSNHFRVERISQEVLAAGLEFGNRMLEKSGFPAQRPRLTEVAVRKTK 128
336	IKIL-----AYDETTOKKSASTEGILNQO-----ISCNPMK 367
129	DRRSFYVIFPDGTFPHGPGSQRMLSY-----TANRREYAL----- 168
368	EQ-----CTYPAISFSM-----NQHFLLFCGPRVPVVSLSHTDNPAKTYFLBESNMUK 417
169	---LHPEPRPVLVCHGTEMGRAPDLDAVFAFMKLDLDELGINIVAPVLPNMGPGQGLP 225
418	EALIKKKIKGPEIKIHLIDY-ELPLQISLPRQFMDRQYALLIMD-----EPPGQLVLT 472
226	-----KGAVPRGEDVDVCH-----GTQDAVVDIIRL-----L 253
473	DKFHLDWDSVLIDMDNVIVARFDGSGSGQGLKILQEIHRIRGSVEVVDQITAVVFLTK 532

```

Qy 254 SWIRSQEESLIGLNGLSLGGYIASLVASLEEG-ACAILGVPAVL 299
      : : : : : : : : : : : : : : : : : : : : : :
Db 533 PYIDSKR-----LSIFGKGYGGYIASMLKSDKELFRGGSVAPITDL 575

```

Search completed: March 23, 2006, 05:10:56  
Job time : 111.803 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 05:11:18 ; Search time 16.4313 Seconds  
(without alignments)  
2418.401 Million cell updates/sec

Title: US-10-617-038-29

Perfect score: 2195  
Sequence: 1 MASSASDGTHERSAPRLSP.....AALBOSGLDAPRTORDRSA 413

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2195	100.0	413	2 B70573	hypothetical prote
2	293.5	13.4	400	2 C70837	hypothetical prote
3	126	5.7	582	2 D72636	probable acylamino
4	114	5.2	741	2 UC5142	X-Pro dipeptidyl-p
5	109	5.0	314	2 E72287	hypothetical prote
6	106.5	4.9	389	2 T30216	hypothetical prote
7	105	4.8	330	2 D97235	probable hydrolase
8	104	4.7	222	2 AB2633	conserved hypotet
9	101.5	4.6	285	2 H72300	hypothetical prote
10	101.5	4.6	467	2 C72601	probable tryptoph
11	99.5	4.5	1303	2 C87519	hypothetical prote
12	97.5	4.4	318	2 B90458	hypothetical prote
13	97	4.4	1020	2 AB0701	tetrathionate redu
14	96	4.4	505	2 F71404	hypothetical prote
15	95	4.3	260	2 H84060	lysophospholipase
16	95	4.3	305	2 F69966	hypothetical prote
17	95	4.3	890	2 F75289	alanyl-tRNA synth
18	94	4.3	475	2 E90318	medium-chain-fatty
19	94	4.3	2082	2 T37056	probable multi-dom
20	93.5	4.3	359	2 B64919	conserved hypotet
21	93.5	4.3	359	2 D90920	hypothetical prote
22	93.5	4.3	359	2 AB5769	hypothetical prote
23	93.5	4.3	1286	1 RJ0BP	interphotoreceptor
24	93.5	4.3	1668	1 A60272	IGA-specific metal
25	93.5	4.3	2324	1 A29924	acetyl-CoA carboxy
26	92.5	4.2	478	2 B72543	hypothetical prote
27	92.5	4.2	1212	2 F83153	probable two-compo
28	92	4.2	361	2 H75260	conserved hypotet
29	91.5	4.2	391	2 A97633	hypothetical prote

30	91.5	4.2	391	2 AD2856	conserved hypotet
31	91.5	4.2	1660	2 T17485	peptide synthetase
32	91	4.1	421	2 TC7367	second peroxidomal
33	91	4.1	633	2 T17262	hypothetical prote
34	91	4.1	1159	2 E83237	probable two-compo
35	90.5	4.1	307	2 C85363	hypothetical prote
36	90.5	4.1	502	2 S30916	bchx protein - Rho
37	90.5	4.1	502	2 T50754	prochlorophyllid
38	90.5	4.1	505	2 T03394	probable betaine-a
39	90.5	4.1	925	2 S27920	nuclear antigen EB
40	90	4.1	671	2 C81033	prolyl oligopeptid
41	89.5	4.1	400	2 H82833	proline dipeptid
42	89.5	4.1	570	2 D72597	hypothetical prote
43	89.5	4.1	593	1 A69655	two-component sens
44	89	4.1	357	2 C72746	probable carbamoyl
45	89	4.1	606	2 T35378	probable peptide h

ALIGNMENTS

RESULT 1

B70573  
hypothetical protein Rv2627c - Mycobacterium tuberculosis (strain H37RV)

C/Spectes: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004

C/Accession: B70573

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felcwell, T.; Gentile, S.; Hamlin, N.; Holtrold, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295387; PMID:9634230

A/Accession: B70573

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-413 <COL>

A/Cross-references: UNIPROT:006185; UNIPARC:UP100000CCAF2; GB:295387; GB:AL123456; NID:9

A/Experimental source: strain H37RV

C/Genetics:

A/Genes: Rv2627c

Query Match	Score	Length	ID	Description
Best Local Similarity	100.0%	413		
Matches	413	Conservative	0	Mismatches 0; Indels 0; Gaps 0;
QY	1	MASSASDGTHERSAPRLSPVLSGAMGPPMTGLVYVQSWRDYLCQOPDKLPAPRTIAL	60	
DB	1	MASSASDGTHERSAPRLSPVLSGAMGPPMTGLVYVQSWRDYLCQOPDKLPAPRTIAL	60	
QY	61	MAQAFRDELIVLGLKARPPVSNHRVFERISQVVAAGLEFYGNRRMLKXSGSFAPQPPPT	120	
DB	61	MAQAFRDELIVLGLKARPPVSNHRVFERISQVVAAGLEFYGNRRMLKXSGSFAPQPPPT	120	
QY	121	EVAVRKVKDRRRSFYRIFPDGFTPHGPGSQRMSTYANRREVALLRHSEPPRWLVC	180	
DB	121	EVAVRKVKDRRRSFYRIFPDGFTPHGPGSQRMSTYANRREVALLRHSEPPRWLVC	180	
QY	181	VHGTMGRAPLDLAVFRANKLHDELGLNIVMVLPMHGRGGCLPKGAVFPGEVDLDVH	240	
DB	181	VHGTMGRAPLDLAVFRANKLHDELGLNIVMVLPMHGRGGCLPKGAVFPGEVDLDVH	240	
QY	241	GTAQAVMDIRRLLSMIRSGEESLGLGLGLSGVYIASIVASLEGICALITGVPAADI	300	
DB	241	GTAQAVMDIRRLLSMIRSGEESLGLGLGLSGVYIASIVASLEGICALITGVPAADI	300	
QY	301	ELLGRHCGLRHNDPRRHVTKAEPFGIMISPLSLPLVPMRFTYAGIADRLVHPRQV	360	
DB	301	ELLGRHCGLRHNDPRRHVTKAEPFGIMISPLSLPLVPMRFTYAGIADRLVHPRQV	360	
QY	361	TLWMHMGKEIVWVPGHGTGFQSGRPVRRVQAALBOSGLDAPRTORDRSA	413	
DB	361	TLWMHMGKEIVWVPGHGTGFQSGRPVRRVQAALBOSGLDAPRTORDRSA	413	





QY 213 -----VLPNHGPRGQ-----GLP-KGAV 229  
 DB 510 GEDPAKRYVAAVYVGGQVITQDTPSGRGDHLFNQYLAQOQVVPVSLDNRTGTPRGRD 569  
 QY 230 PGEEDVLDVHGTAGVAVMDIRRLLSWIRSOE--EESLIGLNGSLGTYIA-SLVALEEG 286  
 DB 570 FGG--ALYKQGVTE-VADQLRGVAVLKKQPPWDPRARIGVQKSGNGVYTMILAAASD 626  
 QY 287 LACAILGVFVADLIELGRHCGLRHND-PRRHTVKAEP-----IGRMISPLSLTPVPM 340  
 DB 627 YACGVAGAVPTDW-GLYDSHYTERVMDLPRANDAGREARVLTHIGLRSPLE----- 677  
 QY 341 PGRFYTAGIADR-LVHPREQVRLMEHNGKPEIYVPGSGHTGFQSPRPVRRF 391  
 DB 678 ---LTHGMADNVLTFTNSTLSMALQKGGPELMTYPGARKGLSGDALHRY 728

## RESULT 5

E72287  
 hypothetical protein - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004

C:Accession: E72287

R.Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.

C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq

A:Reference number: A72200; PMID:99287316; PMID:10360571

A:Accession: E72287

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-314 <ARN>

A:Cross-references: UNIPROT:Q9X0P3; UNIPARC:UP100000C12B6; GB:AE001773; GB:AE000512; NID

C:Experimental source: strain MSB8

C:Genetic8:

A:Gene: TM1159

Query Match 5.0%; Score 109; DB 2; Length 314;

Best Local Similarity 23.7%; Pred. No. 0.19; Mismatches 97; Indels 65; Gaps 14;

Matches 61; Conservative 34; Mismatches 97; Indels 65; Gaps 14;

QY 195 VERAWKLI--DELGLNIYVPLPMHGPRGQGLPKGAVFGEEDVLDVHGTAGVAV----- 247  
 DB 70 LYLIMAAHYLSRGIRAILPVLPGNFTR--VAEGSV-SGKDYFSSDLRMSRFWEHALT 125  
 QY 246 DIRRLLSWIRSOE--EESLIGLNGSLGTYIASLVALEEGACAILGVPAVDLIEL- 303  
 DB 126 DLJLLELLKVKMKMHERNC--LFGYCLGSMIAVLNLALSDPKTTIMAGDEPATLFW 183  
 QY 304 -----GRHGLRHK-----DPRRHVTK-----MAEPIGRMISP-LSLT 335  
 DB 184 KSPITLSFVRRLKSGGSEHGHTNENFYDVRSDLERLKEPSSVQEMSSNHLPLKLD 243  
 QY 336 PLVMPGREFYAGIAD-RLVH-----PREQVRLMEHNGKPEIYVPGSGHTGF-- 382  
 DB 244 PLA-----YAKFVDTSRIVMLEAWFDRALPKSTRDILMEHLGKXRIKVPSSHVSMLP 296  
 QY 383 FOSRPVRRFVQAALEQS 399  
 DB 297 FQVLARAYIMKLIKEQS 313

## RESULT 6

T30216

hypothetical protein Z - Streptomyces hygroscopicus

C:Species: Streptomyces hygroscopicus

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T30216

R.Aparticio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Stuart

Gene 169, 9-16, 1996

A:Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg

A:Reference number: Z20782; PMID:96186896; PMID:8635756

A:Accession: T30216  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-389 <APA>  
 A:Cross-references: UNIPROT:Q54287; UNIPARC:UP100000BE1FA; EMBL:X86780; NID:9987088; PID  
 C:Superfamily: Streptomyces hygroscopicus hypothetical protein Z

Query Match 4.9%; Score 106.5; DB 2; Length 389;

Best Local Similarity 24.1%; Pred. No. 0.4;

Matches 71; Conservative 25; Mismatches 122; Indels 77; Gaps 13;

QY 113 FAQPPRLTEAVARXK-----DRRRSFYRIFFDSGFTPHGSEBSQRLSYTANNEVALL 168  
 DB 112 FDDAARPLDTPRAAIRIPYEDTLPGYLFVDSGTPR----- 150  
 QY 169 LNHPRPRLVCVHGTENGARPLDLAVFRAMKDELGLNIYVPLPMHGPRGQ---LP 225  
 DB 151 -----TVLFHGQYSTLEEDYLAALAGL--RGRV-----VIAFDGP-GQGSYARE 194  
 QY 226 KCAVFPGEEDVLDVHGTAGVAVMDIRRLLSWIRSOEESLIGLNGSLGTYIASLVALE 285  
 DB 195 QGHNFP-----PDMAVAVTPVVDFAITLTPREV--DAEQLV-LMGKSLGTYLAARAAPERH 245  
 QY 286 GLACAILGVPAVDLIELGRHCGLRHNDPRRHTVMAEPI-----GRMISPLS-LTP 336  
 DB 246 RVAAACVLVDGYVDFHVAIVAAITAGRAASTPGGMEALMAQNTVVRVVRNGRWTGVSQVDE 305  
 QY 337 LVPMPGREFYAGIADRLVHPREQVRLMEHNGKPEIYVPGSGHTGFQSPRPVRRF 391  
 DB 306 LVRAAEPTTMAGIARITCP-----TLVLE-----AENDGFFGQGPRLF 345

## RESULT 7

D97235

probable hydrolase from alpha/beta family, YOKD B. subtilis ortholog [imported] - Clostr

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 05-Oct-2004

C:Accession: D97235

R.Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

J.; Daly, M.D.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A:Reference number: A96900; PMID:21359325; PMID:21359325

A:Accession: D97235

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-330 <KUR>

A:Cross-references: UNIPROT:Q97FLL; UNIPARC:UP100000CA5FA; GB:AE001437; PIDN:AAK80671.1;

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetic8:

A:Gene: CAC2725

C:Superfamily: Alpha/beta hydrolase

Query Match 4.8%; Score 105; DB 2; Length 330;

Best Local Similarity 21.1%; Pred. No. 0.43; Mismatches 82; Indels 46; Gaps 8;

Matches 44; Conservative 37; Mismatches 82; Indels 46; Gaps 8;

QY 173 EPRPWLVCVHGTENGARPLDLAVFRAMK-----LHDEGLNIYVPLPMHGPRGQGLPK 226  
 DB 89 DSRKTIIVHGTGSR-----WESMKYADIIYLDGIVNVLVYDSRIHVGSG----- 133  
 QY 227 GAVFPGEEDVLDVHGTAGVAVMDIRRLLSWIRSOEESLIGLNGSLGTYIASLVALE 285  
 DB 134 -----GNDI---TLGYFER-YDLANNVAVKVKKTGPGIIGIHGSEMGATATLQGNMKE 184  
 QY 286 --GLACAILGVPAVDLIELGRHCGLRHNDPRRHTVMAEPIGRM-----ISF 331  
 DB 185 TKDVFVYVDDCFPSDLPLDFGEKLVNEMKNHGAIVAKVAVFYSSLIAEFKAGFSVYALSP 244  
 QY 332 LSLTLPVMPGREFYAGIADRLVHPREQV 360  
 DB 245 IKAIDQVKTPIWFAH-GADDDLIIPENSV 272

**RESULT 8**

A62533 conserved hypothetical protein XF1929 [imported] - Xylella fastidiosa (strain 9a5c)  
C|Species: Xylella fastidiosa  
C|Date: 18-Aug-2000 #sequence\_rev10n 20-Aug-2000 #text\_change 05-Oct-2004  
C|Accession: A62533  
R|Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen-

Nature 406, 151-157, 2000  
A|Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A|Reference number: A62515; MUID:20365717; PMID:10910347  
A|Note: for a complete list of authors see reference number A59328 below  
A|Accession: A62533  
A|Status: preliminary  
A|Molecule type: DNA  
A|Cross-references: UNIPROT:Q9PCF2; UNIPARC:UPI00000C284A; GB:AEO04004; GB:AEO03849; NITE  
A|Experimental source: strain 9a5c  
B|Strain, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, B.  
de-Nero, E.; Docena, C.; El-Dorriy, H.; Facinani, A.F.; Ferreira, A.J.S.

Submitted to GenBank, June 2000  
A|Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm,  
J.D.; Junqueira, M.L.; Kemper, E.L.; Ktajiama, J.P.; Krieger, J.B.; Kuramae, E.E.; Laigret,  
chado, M.A.; Medeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins,  
Rodrigues, V.; Rosa, A.J. de M., de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasaka,  
A|Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira,  
M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A|Reference number: A59328  
A|Contents: annotation  
C|Genetics:  
A|Gene: XF1929  
C|Superfamily: uncharacterized conserved protein

**Query Match** 4.7% Score 104; DB 2; Length 222;  
**Best Local Similarity** 25.0%; Pred. No. 0.31;  
**Matches** 38; Conservative 24; Mismatches 60; Indels 30; Gaps 8;

OY 228 AVPGEDVLDVDHGTGAQAVMDIRLLSWIRSQEESLIGLNGLSIGYIASVASLEEG 287  
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
DB 81 SVGSADGMPDGNHGGERA---DLRAIAAWVRAPDPSPALMLAGFSGAYISLVAAEELEPQ 137  
OY 288 ACATLGVAVALIELLGHCHGRKHDPRRRHYKMAEPFGIMISP-----LSLP----L 337  
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
DB 138 VLISISPPA-----GR-WDFSHVHPPEHMLLIGDADEVDDPAQVYDWITSLPROPGL 189  
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
OY 338 VPMP--GRFIYAGIADRVLVHPRECV--TRLM 364  
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
DB 190 IRMPDTSHFPFR---KLHLRDALQDGVRSM 217  
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

**RESULT 9**

H72300 hypothetical protein - Thermotoga maritima (strain MS86)  
C|Species: Thermotoga maritima  
C|Date: 11-Jun-1999 #sequence\_rev10n 11-Jun-1999 #text\_change 09-Jul-2004  
C|Accession: H72300  
C|Netlson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
Nature 399, 323-329, 1999  
A|Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A|Reference number: A72200; MUID:99287316; PMID:10360571  
A|Accession: H72300  
A|Status: preliminary  
A|Molecule type: DNA  
A|Residues: 1-285 <ARN>  
A|Cross-references: UNIPROT:Q9XND6; UNIPARC:UPI00000C1315; GB:AEO01765; GB:AEO00512; NITE  
A|Experimental source: strain MS86  
C|Genetics:  
A|Gene: TM1046

	Query Match	4.6%; Score 101.5; DB 2; Length 285;
	Best Local Similarity	21.7%; Pred. No. 69;
	Matches	65; Conservative 44; Mismatches 133; Indels 57; Gaps 15;
Oy	137 IFPDSGFPHGEGESQRMVSYTANNEYALLLHPPEPRMLVCVHSTENGRAPLDAVF 196	
Dd	4 ITPDASVDPIAESEKTRVVYVEBEKERULIF-----LHGIGNGP-----Y 47	
Oy	197 RAW--KLHDEGLNIWMPVLPMGSPRG-QGLPKGAVPFGEVDLDVHGTAQAVMDIRLL 253	:
Dd	48 LLMGEEKFRREYNITKTWFILIPYHEKRAPENWSGGEFPYHSSPSFCVKRPFDAVDIDL 107	:   :
Oy	254 SWMSQEBESLIGNGSLSGYTASLVAS-----LEGLACAILG-----VPVADIE 301	:::::
Dd	108 DVNR-RNDKDFRISLMGSFGSMITTILLAEKRIEKGVICCSGDWMWIMYSFYTERLR 166	:::::
Oy	302 LL-----GRHCGLR-HKD---PRRT-----VKNAEPIC-RLMISPLSTPLVPM 340	:
Dd	167 ELYRKNGNMGECREKCKIKRRARPAELIKSFNSIEDIRKKRPVGCYFYDPASPAPFDVQ 226	:
Oy	341 PGRITVAGIDRLVHPBQVTRLMEHWKCEIYWPCGHNG--FFQSRRPVRRFOALE 397	:
Dd	227 KLFEFMA-LPDHVI-PYQSYACLHKLNKKKTIVLPBGHGSYFFRRYIAKRVVRFLID 283	:
	RESULT 10	
Ct	C72601	
P	Problems: cryptophanase APE1275 - Aeropyrum pernix (strain KI)	
C	Species: Aeropyrum pernix	
C	Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004	
C	Accession: C72601	
R	Rikawabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haitakawa, Y.; Jhn-no, K.; Takahara, H.; Takamaya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kawa, R., 83-101, 1999	
A	Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum	
A	Reference number: A72450; PMID:10382966	
A	Accession: C72601	
A	Status: preliminary	
A	Molecule type: DNA	
A	Residues: 1-467 <KW>	
A	Cross-references: UNIPROT:Q9YC12; UNIPARC:UP100000SDEA6; DDBJ:AP000061; NID:g5104821,	
A	Experimental source: strain KI	
C	Genetics:	
A	Gene: APE1275	
C	Superfamily: tryptophanase	
	Query Match	4.6%; Score 101.5; DB 2; Length 467;
	Best Local Similarity	22.5%; Pred. No. 1.3;
	Matches	75; Conservative 48; Mismatches 106; Indels 105; Gaps 19;
Oy	65 FRDEI-VLLGKARKRPPSNHRVPERI-----SQEVAAGLEFYGNNRWLEKEPGCF 112	
Dd	86 FRDAVDVLGLDLVLPVHQGRPAERILYGBELLRRRNRIVPANTHTPDTGRAVILNQGV 145	
Oy	113 PAOPPLETEVAVRKYKORRSFYRIFPDGFTPHPGC---PGSQRWLSYTNANREVALLL 169	:
Dd	146 LDLESP-----QSRREAY-----PFKGIDVARLERILK--ERSRDVAFIL 185	:   :
Oy	170 RHPEPRMLVCVHSTENGRAPLDI-AVFRMKLHDELGLNIWMPV-----LPMHCP 219	:
Dd	186 -----LVTTNTTAGGG--FVSMDNKTVRELDAAGLPLVMDICFAENALYLKERDP 236	:
Oy	220 RGQQLPKGANVPGEDVDVHGTQAAMVDI-----RLLSW-----IRSQEBESLIGNLGS 271	:
Dd	237 RYRG-----WSVVDIAREMTSYGDHFVMSAKDGL-----AN 268	:
Oy	272 LGGYIASLVASLEGGLACAIL-----GVPAVDILELGRHCGLR---HKDPRRTV 319	:
Dd	269 ICGTIARRDSLSYGLDLAARVYLEGCVTYTGGLAGRDL-EALAQ--GLREVVEDYLHRY 325	:
Oy	320 KMAEPIGMISPLSTPLVPMGFRFYAGIADRLL 353	:

Db 326 EQVRYKGLLSQGVIVEPVGGHAYVDVLEAL 359

## RESULT 11

C87519 hypothetical protein CC2180 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #ext\_change 09-Jul-2004

C:Accession: C87519

R:Niemman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.B.; Eiken, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gilm, M.L.; Haft, D.H.; Kolon

n, J.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: C87519

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1303 <STO>

A:Cross-references: UNIPROT:Q9A6B4; UNIPARC:UPI00000C762F; GB:AE005673; NID:g13423679; H

C:Genetics:

A:Gene: CC2180

Query Match 4.5%; Score 99.5; DB 2; Length 1303;  
Best Local Similarity 21.2%; Pred. No. 7.2;  
Matches 89; Conservative 54; Mismatches 138; Indels 139; Gaps 23;

QY 1 MASSASDGTHERSAPFLSPVLSGAMPFMTGLVYAQ-----SMRDYLGQDPDKP 52

DB 422 VASRAKD--QORSVASLDQPLKGGVL-----FTTVAQETPIQELWAYVFGAEPK--D 472

QY 53 IARPTALAAQAFRDSIVLGLKARPVSNHFRFESIQEVAAGLEFGYGNRW-LEKPEG 111

DB 473 ARKMSYTVAAEYAPDVLNADLN-----YLAGRPMEKRAIVVALPFG 516

QY 112 PFAQPPLETVAVRKVKDRRSFYRIEFDGFTPH-GEPSGQRMVSYANNREVALLR 170

DB 517 AENRPAPGDLAAKGL-----PIVHTLIPSGFDAPAAKPLARNW----- 556

QY 171 HEPPRWLVCHGTENGARPLDLAVFRANKHDELGLNIVMPRLPHNGRGGLP----- 225

DB 557 -----AYGWEVARQGLD-----GVALDIPLP--GAAGAVIPLNIQV 591

QY 226 KCAVFPGEVDLD---DVHGTAAVW-DIR-RLSWIRSGEESLIGLNGLSGYIASL 279

DB 592 KQPIPGRMIDVSYSVKAQEARTWLDLRDLIT-----ADSL-----YLTIA 635

QY 280 VASLEBG--LACALIGEVADLIELGRHCGLRHNDP-----RHTVKAEPFGM 328

DB 636 ADSAEGPRALDQAKRLVLKORDQAKVHIQDRLNQVDNNAFLVEHTTSKQGLYRR 695

QY 329 ISPLSLTLPVPMGRFIYAGIDL-VHREOVTRLMHWKPEIYWGCHTGFQSQRP 387

DB 696 -----LYADISDLRLVDPDNKIGR--EYWG--DIYSGQGWPAFEPQRP 735

## RESULT 12

B90458

hypothetical protein SSO2810 [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C>Date: 24-May-2001 #sequence\_revision 24-May-2001 #ext\_change 09-Jul-2004

C:Accession: B90458

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-

g, J.; Ueffler, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thl-NGOC, H.P.; Redder, F

aretz, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.

submitted to Genbank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: B90458

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-318 <KUR>

A:Cross-references: UNIPROT:Q9YV26; UNIPARC:UPI0000064872; GB:AE006641; NID:g13816160; H

C:Genetics:  
A:Gene: SSO2810

Query Match 4.4%; Score 97.5; DB 2; Length 318;  
Best Local Similarity 21.4%; Pred. No. 1.7; Indels 87; Gaps 10;  
Matches 54; Conservative 32; Mismatches 79; Indels 87; Gaps 10;

QY 185 ENKRAPLDL-AVFRAMKLHDELG-----INTVPLPMHGRPGQLEPKAVFGEVDL 236

DB 103 EGYTAPIDQGFALIMWKVDEKVAAYKFGDNLVKKVPIVFPFGVIGTGLP----- 151

QY 237 DVHGTAAQWDIRRLSLWSIQEESLIGLNGLSG--YIASVASLEBLACAILGV 294

DB 152 -----FRKSLSTIPRENGGMMDKHLTVGKYLPEVNG-----ALISI 192

QY 295 PVADLIELGRHCGLRHNDPFRHTVKAEPFGMSPLST-PLVPMGRFIYAGIADRL 353

DB 193 GDTLAAQDGEVCGTAIEHMEVTVK-----RLKNVGLTQPLP----- 232

QY 354 VHPREOVTRLMHWKPEIYWGCHTGFQSQ-----RPVRPVQA-- 395

DB 233 -----VTKVKEMEYDEVIAVPGIDSNLMNATKKAIGTISLSKMAPVEAYILASV 286

QY 396 --LEQSGLIDAP 405

DB 287 VMLRVSEVDVP 298

## RESULT 13

AE0701

tetrachionate reductase chain A [imported] - Salmonella enterica subsp. enterica serovar

C:Species: Salmonella enterica subsp. enterica serovar Typh

A:Note: this species has also been called Salmonella typh

C:Accession: AE0701

C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #ext\_change 18-Nov-2002

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moulie, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Barry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skellon, J.; Stevens, K.;

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AE0502; MUID:21534947; PMID:11677608

A:Accession: AE0701

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1020 <PAR>

A:Cross-references: UNIPARC:UPI0000059DE8; GB:AL513382; PIND:CAD01982.1; PID:g16502823;

C:Genetics:

A:Gene: ttrA

Query Match 4.4%; Score 97; DB 2; Length 1020;  
Best Local Similarity 21.7%; Pred. No. 8.5; Indels 144; Gaps 19;  
Matches 87; Conservative 46; Mismatches 124; Indels 144; Gaps 19;

QY 22 ISGAMPFMTGLVYAQ-----WRDYLGQ-QPDKLPIAPPTIAL-AAQAFRDEIVLLG 73

DB 530 LSG--GVFVGKKEFGVSDGRYNNNSFAGKYKPGSLIASKTAVKXSEERDKIA--G 585

QY 74 LKARPPV--NHRVERISQEVAAAGLEPYGN--RRMLEKSGSFPAPPPLETVAVRKVK 128

DB 586 GQSPVPAPAPWYPFVAGQLTEILTSALBGYPPLKAMISNNSNPYGVPGLRVAVEEKLK 645

QY 129 DRRRSFYIIFPS-----GTPHGEPSQRMVSTANNREYA 166

DB 646 DRRRLPLFLIADAFNMTTALADYVDPDTHNFBESGFAPMGVASK-----A 693

QY 167 LILRHPEDRPMLVCHGTENGARPLDLAVF--RAWKHLDELGLNIVMPVPMHGRPGQ 223

DB 694 TTAAPVAVP--AHHRTADQ-PVSMKAFCAVAKRLH-----LPGFEDRAIT 738

QY 224 LPKGAVPFGEVDLDVHGTAAQWDIRRLSLWSIQEESLIGLNGLSGYSGLGYIASL 283

DB 739 DPGQNTFPLNRAEDFYLVRVAANIAMKGTVPVALANQDILSLGVSR----- 785

Qy 284 EBGACALIGVAVDLIELLGRHCGLRHKDPRRHVTKMAEPIGRMISPLSLPLVMPGR 343  
Db 786 -----LPAI-----QHTLK-ADEVGRV-----A 802  
Qy 344 FTY-----AGIADLVHPREQVTRLMEHWKPEIWM 374  
Db 803 FTYSGRGAPRPPDSGYTEQ-----RLGNAMKKPLQIWM 834

## RESULT 14

F71404  
Hypothetical protein - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
A:Variety: Columbia  
C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 09-Jul-2004  
C:Accession: F71404  
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk  
P.; Medler, H.; Medler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel  
avanagh, T.; Hempel, S.; Kotter, P.; Ertlan, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.  
Nature 391, 485-488, 1998  
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech  
erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; And  
C.; Chalwatzis, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal  
A:Reference number: A71400; MUID:98121113; PMID:9461215  
A:Accession: F71404  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-505 <BEV>  
A:Cross-References: UNIPROT:O23287, UNIPARC:UPI00000A98D0, GB:Z9735, NID:G2244747, PTD:  
C:Genetics:  
A:Map position: 4COP9-4G3845

Query Match 4.4%; Score 96; DB 2; Length 505;

Best Local Similarity 27.5%; Pred. No. 4.2;

Matches 39; Conservative 20; Mismatches 53; Indels 30; Gaps 7;

Qy 169 LHHPRPWLVCV---HGTMRAPLDLAVFRAMKLDLGLNIWVPLPMHGRGQGLP 225  
Db 105 VERPPDRP-LPCVICHGSGGRADASEAIVL-----LPSNITITFLDF---SGSGLS 154  
Qy 226 KGAVPGEVDLDVHGTQAQV---DIRLLSWIRSQEBSLIGLNGSLGCIYASLVA 281  
Db 155 GGE-----HYTLG--WNEKXDLKAVVEYLRTDGNVSLIGLGRSGAVTSLMYG 201  
Qy 282 SLBGLACALIGVAVDLIEL 303  
Db 202 AEDPSIAAVLDSPFSDLVLM 223

## RESULT 15

H84060  
Lysophospholipase BH3288 (imported) - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 12-Jul-2004

C:Accession: H84060  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: H84060  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-260 <STO>

A:Cross-References: UNIPROT:Q9K7S1, UNIPARC:UPI00000C415D, GB:AP01518, GB:BA000004; NID:  
A:Experimental source: strain C-125

C:Genetics:  
A:Gene: BH3288

C:Superfamily: trophinesterase

Query Match 4.3%; Score 95; DB 2; Length 260;  
Best Local Similarity 25.5%; Pred. No. 2.1;

Matches 56; Conservative 27; Mismatches 69; Indels 68; Gaps 12;  
Qy 173 EPRPWLVCVHGT--EMGRAPLDLAVFRAM--KLHDELGLNIWVPLPMHGRGQGLPKGA 228  
Db 8 EPRGVVVVHHGAGEHGR-----YQWLAKKFNISGLSVVMDLP-----GQORTGK 54  
Qy 229 VFPGEVDLDVHGTQAQVMD--IRLLSWIRSQEBSL--IGLNGSLGCIYASLVALEEG 286  
Db 55 -----RGHIQSFQOYIDVLEBWEAKLEHVPILFLGHSMGGLVA--VRTMIEG 101  
Qy 287 ---LACALIGVAVDLIELLGRHCGLRHKDPRRHVTKMAEPIGRMISPLSLPLVMPGR 342  
Db 102 GTLPVRAVILSSPCEDLYQSPGKGLASKMLHKTTPFSH----- 142  
Qy 343 RFIYAGIADLVHPREQVTRLMEHWKPEI-----VMY 375  
Db 143 ---HSGIRSDLVTRNEIR---EAYLDELAVTKVSTKMV 176

Search completed: March 23, 2006, 05:26:36  
Job time : 17.4313 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 04:56:43 ; Search time 105.63 Seconds  
(without alignments)  
2758.529 Million cell updates/sec

Title: US-10-617-038-29

Perfect score: 2195  
Sequence: 1 MASSASDGTHERSAPRLSP.....AALEOSGLDAPRTQDRSA 413

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues  
Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: uniprot\_05.80:\*  
2: uniprot\_sprot:\*  
3: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2195	100.0	413	2	006185 MYCTU
2	2188	99.7	413	2	07DEV3 MYCTU
3	2188	99.7	413	2	07MY71 MYCBO
4	328.5	15.0	373	2	073KD4 MYCPA
5	295.5	13.5	339	2	073TC9 MYCPA
6	293.5	13.4	400	2	08VKN4 MYCTU
7	293.5	13.4	400	2	053697 MYCTU
8	293.5	13.4	400	2	07UC3 MYCBO
9	208.5	9.5	325	2	065172 BACID
10	178	8.1	333	2	081CMI BACR
11	158	7.2	326	2	084FBI MYXA
12	153.5	7.0	375	2	09CTC5 ARATH
13	148	6.7	335	2	09LH51 ARATH
14	148	6.7	360	2	09CTD8 ARATH
15	148	6.7	363	2	08RMA3 ARATH
16	146.5	6.7	366	2	06H7Q9 ORYSA
17	139.5	6.4	615	2	07NNY4 GLOVI
18	132	6.0	359	2	04NZ14 GDELT
19	126	5.7	582	1	APBH ABRPE
20	121	5.5	338	2	076462 DROME
21	118	5.4	745	2	06F317 GPRSD
22	117.5	5.4	248	2	064UC0 BACR
23	116.5	5.3	368	2	06C7Y6 YARLI
24	116.5	5.3	574	2	073S12 MYCPA
25	115.5	5.3	248	2	05LD97 BACFN
26	114	5.2	741	2	P95782 XANMA
27	114	5.2	1868	2	0939Y9 GPRSU
28	112.5	5.1	1322	2	08PR21 XANAC
29	112	5.1	832	2	098G05 RHIL0
30	111	5.1	222	2	087CM7 XYLEP
31	110.5	5.0	1925	2	05J106 YNOCA

32	110	5.0	310	2	Q5W1X0 BACSK	Q5W1X0 bacillus c1
33	109	5.0	314	2	Q9X0P3 THEMA	Q9X0P3 thermocoga
34	109	5.0	502	2	Q8GWC9 ARATH	Q8GWC9 arabidopsis
35	108.5	4.9	324	2	Q6F174 MESFL	Q6F174 mesoplasma
36	108.5	4.9	751	2	Q4UPD3 XANCP	Q4UPD3 xanthomona
37	108.5	4.9	751	2	Q8P3V8 XANCP	Q8P3V8 xanthomona
38	108	4.9	338	2	Q84H72 BURMA	Q84H72 burkholderi
39	108	4.9	784	2	Q604E5 METCA	Q604E5 methylococ
40	107	4.9	374	2	Q721H2 THEHT2	Q721H2 thermus the
41	107	4.9	389	2	Q8PFD7 XANAC	Q8PFD7 xanthomona
42	106.5	4.9	389	2	Q54287 STRHY	Q54287 streptomyce
43	106.5	4.9	731	1	GLGB BORPA	Q7614 bordetella
44	106.5	4.9	749	2	Q7NEK8 GLOVI	Q7NEK8 gloeobacter
45	106.5	4.9	7756	2	Q701M5 BREBE	Q701M5 brevivibacill

ALIGNMENTS

RESULT 1	006185 MYCTU	PRELIMINARY, PRT, 413 AA.
ID	006185 MYCTU	
AC	006185 MYCTU	
DT	01-JUL-1997 (TRIMBLrel. 04, Created)	
DT	01-AUG-1998 (TRIMBLrel. 07, Last sequence update)	
DT	01-OCT-2003 (TRIMBLrel. 25, Last annotation update)	
DE	Hypochemical protein.	
GN	OrderedlocusNames=Rv2627C;	
OS	Mycobacterium tuberculosis.	
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium;	
OC	Mycobacterium tuberculosis complex.	
OX	NCBI_Taxid=1773;	
RN	[1]	
RP	NUCLEOTIDE SEQUENCE.	
RC	STRAIN=H37RV;	
RC	MEDLINE=98295987; PubMed=9634230; DOI=10.1038/31159;	
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C.M.,	
RA	Harris D.E., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III,	
RA	Tekala F., Badcock K., Basham D., Brown D., Chillingworth T.,	
RA	Comor R., Davies R.M., Devlin K., Feltwell T., Gentles S., Hamlin N.,	
RA	Holroyd S., Hornby T., Jagels K., Krogh A., McLean J., Moule S.,	
RA	Murphy J.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,	
RA	Rogers J., Rutter S., Seeger K., Skellon S., Squares R.,	
RA	Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;	
RT	"Deciphering the biology of Mycobacterium tuberculosis from the	
RT	complete genome sequence."	
RL	Nature 393:537-544(1998).	
DR	EMBL; BX84580; CAB0637.1; -; Genomic_DNA.	
DR	PIR; B70573; B70573.	
DR	TubercuList; Rv2627C; -.	
DR	GO; GO:0016787; F:hydrolyase activity; IEA.	
DR	InterPro; IPR000379; Ser_Peptide.	
KW	Complete proteome; Hydrolyase.	
SQ	SEQUENCE 413 AA; 46252 MW; 760893D0508523B0 CRC64;	
Query Match	100.0%; Score 2195; DB 2; Length 413;	
Best Local Similarity	100.0%; Pred. No. 1e-158;	
Matches 413; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MASSASDGTHERSAPRLSPVLGAMGPFMTGLVVAQSWRDYLGQDPKLPAPLAL 60	
DB	1 MASSASDGTHERSAPRLSPVLGAMGPFMTGLVVAQSWRDYLGQDPKLPAPLAL 60	
QY	61 AAQAFDEIVLGLKARPPVSNHRVFERISQVAAAGLEFYGNRWLEKPSGFFAPPLT 120	
DB	61 AAQAFDEIVLGLKARPPVSNHRVFERISQVAAAGLEFYGNRWLEKPSGFFAPPLT 120	
QY	121 EAVAVKVDORRSFRIFPDGTFPHGPGSGQRLSTANNREYALLRHBPAPPLVC 180	
DB	121 EAVAVKVDORRSFRIFPDGTFPHGPGSGQRLSTANNREYALLRHBPAPPLVC 180	
QY	181 VAGTEWGAFLDLAVFRAMKLDLGNLIVMVLPMHGPBGGLPKGAVFGEDVLDVHV 240	

Db 181 VGTGEMGRAPLDIAVRAWMLHDELGLNIVMPVLPWHPGPGQGLPGAVFPGEVDLDDVH 240  
 Qy 241 GTAQAVMDIRRLLSWIRSQEESLIGLNGLSIGYIASLVASLEBGLACAILGVPAADI 300  
 Db 241 GTAQAVMDIRRLLSWIRSQEESLIGLNGLSIGYIASLVASLEBGLACAILGVPAADI 300  
 Qy 301 ELLGRHCGLRHKDPRRHVYKMAEPGRMISPLSLTLPVMPGGRFYAGIADRLVHPRQV 360  
 Db 301 ELLGRHCGLRHKDPRRHVYKMAEPGRMISPLSLTLPVMPGGRFYAGIADRLVHPRQV 360  
 Qy 361 TRLMHWKGPETIWMYPGHGTGFPQSRPVRRFVQALLESGLLDAPRTORDRSA 413  
 Db 361 TRLMHWKGPETIWMYPGHGTGFPQSRPVRRFVQALLESGLLDAPRTORDRSA 413

## RESULT 2

Q7D6V3\_MYCTU PRELIMINARY; PRT; 413 AA.  
 AC Q7D6V3;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Hypothetical protein M2660c;  
 GN OrderedLocustNames=M2660c;  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;  
 OC Mycobacterium tuberculosis complex.  
 OK NCBI\_TaxId=1773;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RX MEDLINE=2206494; PubMed=12218036;  
 RA DOI=10.1126/JB.184.19.5479-5490.2002;  
 RA Fietischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 Peterson J.D., Deboy R.T., Dodson R.J., Gilm M.L., Hatt D.H.,  
 Hickey B.K., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D.,  
 Salberg S.L., Delcher A., Utterback T.R., Weidman J.F., Kouri H.M.,  
 Gill J., Mink A., Bishai W., Jacobs W.R. Jr., Venter J.C.,  
 Fraser C.M.;  
 RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RT J. Bacteriol. 184:5479-5490(2002).  
 DR EMBL, AE000516; AAK47018.1; -; Genomic\_DNA.  
 DR TIGR, MT2702; -.  
 DR GO: GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro: IPR000379; Ser\_estrs.  
 KW Hydrolase; Hypothetical protein.  
 SQ SEQUENCE 413 AA; 46153 MW; 7305EBA059FA8F90 CRC64;

Query Match 99.7%; Score 2188; DB 2; Length 413;  
 Best Local Similarity 99.8%; Pred. No. 3.5e-158;  
 Matches 412; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MASSASDGTHERSAPRLSPVLSGAMGPFMTGLVYAQSWRDYLGQOPDKLPAPRTIAL 60  
 Db 1 MASSASDGTHERSAPRLSPVLSGAMGPFMTGLVYAQSWRDYLGQOPDKLPAPRTIAL 60  
 Qy 61 AAQAFRDEIVLLGLKARRPVSNHRRVERISQEVAAGLEFYGNRMWLEKPSGFFAOPPLT 120  
 Db 61 AAQAFRDEIVLLGLKARRPVSNHRRVERISQEVAAGLEFYGNRMWLEKPSGFFAOPPLT 120  
 Qy 121 EVAVKKVDRRRSFRIFPDSGFTPHGEPGSGQRLSTANNREYALLRRHPRPWLVC 180  
 Db 121 EVAVKKVDRRRSFRIFPDSGFTPHGEPGSGQRLSTANNREYALLRRHPRPWLVC 180  
 Qy 181 VGTGEMGRAPLDIAVRAWMLHDELGLNIVMPVLPWHPGPGQGLPGAVFPGEVDLDDVH 240  
 Db 181 VGTGEMGRAPLDIAVRAWMLHDELGLNIVMPVLPWHPGPGQGLPGAVFPGEVDLDDVH 240  
 Qy 241 GTAQAVMDIRRLLSWIRSQEESLIGLNGLSIGYIASLVASLEBGLACAILGVPAADI 300  
 Db 241 GTAQAVMDIRRLLSWIRSQEESLIGLNGLSIGYIASLVASLEBGLACAILGVPAADI 300

Db 241 GTAQAVMDIRRLLSWIRSQEESLIGLNGLSIGYIASLVASLEBGLACAILGVPAADI 300  
 Qy 301 ELLGRHCGLRHKDPRRHVYKMAEPGRMISPLSLTLPVMPGGRFYAGIADRLVHPRQV 360  
 Db 301 ELLGRHCGLRHKDPRRHVYKMAEPGRMISPLSLTLPVMPGGRFYAGIADRLVHPRQV 360  
 Qy 361 TRLMHWKGPETIWMYPGHGTGFPQSRPVRRFVQALLESGLLDAPRTORDRSA 413  
 Db 361 TRLMHWKGPETIWMYPGHGTGFPQSRPVRRFVQALLESGLLDAPRTORDRSA 413

## RESULT 3

Q7TY71\_MYCBO PRELIMINARY; PRT; 413 AA.  
 AC Q7TY71;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Hypothetical protein M2660c.  
 GN OrderedLocustNames=M2660c;  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;  
 OC Mycobacterium tuberculosis complex.  
 OK NCBI\_TaxId=1765;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=AF2122/97;  
 RX MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;  
 RA Garner T., Bigliamer K., Camus J.-C., Medina N., Manoor H.,  
 Pryor M., Duchoy S., Giondi S., Lacroix C., Monsemp C., Simon S.,  
 Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 Parhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
 RA "The complete genome sequence of Mycobacterium bovis."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
 DR EMBL, BX248343; CAD94845.1; -; Genomic DNA.  
 DR GO: GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro: IPR000379; Ser\_estrs.  
 KW Complete proteome; Hydrolase.  
 SQ SEQUENCE 413 AA; 46153 MW; 7305EBA059FA8F90 CRC64;

Query Match 99.7%; Score 2188; DB 2; Length 413;  
 Best Local Similarity 99.8%; Pred. No. 3.5e-158;  
 Matches 412; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MASSASDGTHERSAPRLSPVLSGAMGPFMTGLVYAQSWRDYLGQOPDKLPAPRTIAL 60  
 Db 1 MASSASDGTHERSAPRLSPVLSGAMGPFMTGLVYAQSWRDYLGQOPDKLPAPRTIAL 60  
 Qy 61 AAQAFRDEIVLLGLKARRPVSNHRRVERISQEVAAGLEFYGNRMWLEKPSGFFAOPPLT 120  
 Db 61 AAQAFRDEIVLLGLKARRPVSNHRRVERISQEVAAGLEFYGNRMWLEKPSGFFAOPPLT 120  
 Qy 121 EVAVKKVDRRRSFRIFPDSGFTPHGEPGSGQRLSTANNREYALLRRHPRPWLVC 180  
 Db 121 EVAVKKVDRRRSFRIFPDSGFTPHGEPGSGQRLSTANNREYALLRRHPRPWLVC 180  
 Qy 181 VGTGEMGRAPLDIAVRAWMLHDELGLNIVMPVLPWHPGPGQGLPGAVFPGEVDLDDVH 240  
 Db 181 VGTGEMGRAPLDIAVRAWMLHDELGLNIVMPVLPWHPGPGQGLPGAVFPGEVDLDDVH 240  
 Qy 241 GTAQAVMDIRRLLSWIRSQEESLIGLNGLSIGYIASLVASLEBGLACAILGVPAADI 300  
 Db 241 GTAQAVMDIRRLLSWIRSQEESLIGLNGLSIGYIASLVASLEBGLACAILGVPAADI 300  
 Qy 301 ELLGRHCGLRHKDPRRHVYKMAEPGRMISPLSLTLPVMPGGRFYAGIADRLVHPRQV 360  
 Db 301 ELLGRHCGLRHKDPRRHVYKMAEPGRMISPLSLTLPVMPGGRFYAGIADRLVHPRQV 360  
 Qy 361 TRLMHWKGPETIWMYPGHGTGFPQSRPVRRFVQALLESGLLDAPRTORDRSA 413  
 Db 361 TRLMHWKGPETIWMYPGHGTGFPQSRPVRRFVQALLESGLLDAPRTORDRSA 413



```

RESULT 4
073XD4 MYCPA PRELIMINARY: PRT: 373 AA.
ID 073XD4 MYCPA PRELIMINARY: PRT: 373 AA.
AC 073XD4 MYCPA PRELIMINARY: PRT: 373 AA.
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE Hypochemical protein.
GN OrderedlocusNames=MAP2375C;
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium avium complex (MAC).
NCBI_TaxId=1770;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K10;
RA Li L., Bannantine J., Zhang Q., Amonsin A., Alt D., Kapur V.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; A6017235; AAS04692.1; -; Genomic DNA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR000379; Ser_estrs.
DR Complete proteome; Hydrolase.
KW SEQUENCE 373 AA; 40489 MW; 6477DD7D520E296A CRC64;
SQ
Query Match 15.0%; Score 328.5; DB 2; Length 373;
Best Local Similarity 29.8%; Pred. No. 1.3e-16;
Matches 109; Conservative 46; Mismatches 148; Indels 63; Gaps 13;

QY 44 LGGQPDKLPAPPTTALAAQAFDEIVLLGLKARFVSNHRYEERISQEVAAGLEFYGNR 103
DB 62 LGGSPALL---RPVDCAAA---EELALGI----- 87
QY 104 KMLEKPSGFAAPPTTEVAVKVKRRSPFRIEFD--SGPTPHGSPGSRKSLSTANN 162
DB 88 -----DGHAAAPNPLRPTSI-----RRSISGLDTERMAFEHDPLPTTLVADGIGCPA 136
QY 163 REYALLLRHPE--PRPMLVCVCHGTEMGRAPLDLAVFPAWKLHDELGINIWPVLPMHGPRG 221
DB 137 RAVVHLRRRRDGRPLVWVHAGGQCTE--DLLSGIDRIHHLGFINIMPVOPGHGCR 195
QY 222 QGLPKGAVPEGEVDLDVHGTAAQAVDRIRLLSWISQSEESLIGLNGSLGGYIASLVA 281
DB 196 RQWP---VYPMDDPLGNVAGMRAVSEVAAYVMASSQ--ASTLVAGISMSGSPVALVVS 250
QY 282 SLEEGIAACILGVPAVDLIELGRHCGLRHKDPRHTVR-----MAEPIGMISPLSLT 335
DB 251 HLERRVDAVALYTPITGLNAMIARH--LQWGPARDGPRELLESPPVTLTSTVIDPLAVT 308
QY 336 PLVPPRGRIYAGIDRLVHPRQVRLMEHMKPEIVVYPGHTGFFQSRPVRVQA 395
DB 309 PAPPPERRILVGAHMDRMA--MREPANALODRMA--GQLVWYDSGHVIFSRVQITDRF 366
QY 396 LEOGSL 401
DB 367 LADAVL 372

```

```

RESULT 5
073TC9 MYCPA PRELIMINARY: PRT: 399 AA.
ID 073TC9 MYCPA PRELIMINARY: PRT: 399 AA.
AC 073TC9 MYCPA PRELIMINARY: PRT: 399 AA.
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE Hypochemical protein.
GN OrderedlocusNames=MAP3789C;
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium avium complex (MAC).
NCBI_TaxId=1770;
KW

```

```

RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K10;
RA Li L., Bannantine J., Zhang Q., Amonsin A., Alt D., Kapur V.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; A6017240; AAS06339.1; -; Genomic DNA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR000379; Ser_estrs.
DR Complete proteome; Hydrolase.
KW SEQUENCE 399 AA; 43866 MW; 6B28A680BD2C697C CRC64;
SQ
Query Match 13.5%; Score 295.5; DB 2; Length 399;
Best Local Similarity 30.7%; Pred. No. 4.5e-14;
Matches 101; Conservative 43; Mismatches 144; Indels 41; Gaps 15;

QY 95 AGLFPGNRRLMEKSGFP---AOP-----PLTEVAVKVDRRRSFYRIFPDSGF 143
DB 84 ANLEFYAEIGAAADPDSPFPPTPEPTTSRRASPLAEVWAGTYD-----NLAFDSGF 137
QY 144 TP-HGEPGSPQRLSYTANNREYALLLRHPE--PRPMLVCVCHGTEMGRAPLDLAVFPAWKL 201
DB 138 TAVNPAM--RAWMSGGRNNVVRQAQHWRRHDGPRPTLCVIHGF--LGSSYLANGRFSLPW 194
QY 202 HDLGLNIWVPLVPMHGPRGGLP--KGAVFGEVDLDVHGTAAQAVDIRLLSWIRSQ 259
DB 195 YRAGYDVLMTYLPFHGKRSERFSPSGFYAGGSGFAEAMAQAVIDFSLIDYLRGT 254
QY 260 EESLSLIGNGSLGYSIVASLEGLACALIGVPV---ADLIE-----LIGRHGL 309
DB 255 GVER--IALTGISLGGTSLVASVDRLNAVTPNCEVVTPTATLFBWFPANKLVG--LGL 311
QY 310 RHKDRPRHTVKAEPGIRMIISPLSLTPLVPMGRIYAGIDRLVHPRQVRLMEHMK 369
DB 312 RSSDISR--AQLAAGLA--YHCLTYRPLVAKORMIITGLGRMA--PPEPAKLMQHMDR 367
QY 370 PEIVWYPGGHTGFFQSRPVRVQAALQ 398
DB 368 CALHWPFGSHVLHISQLDYLRMTAFLOQ 396

```

```

RESULT 6
08VKN4 MYCTU PRELIMINARY: PRT: 400 AA.
ID 08VKN4 MYCTU PRELIMINARY: PRT: 400 AA.
AC 08VKN4 MYCTU PRELIMINARY: PRT: 400 AA.
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Hypochemical protein.
GN OrderedlocusNames=MT0306;
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium tuberculosis complex.
NCBI_TaxId=1773;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RX DOI=10.1128/JB.184.19.5479-5490.2002;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.D., Deboy R.T., Dodson R.J., Gwinn M.L., Haft D.H.,
RA Hickey E.K., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D.,
RA Salzberg S.L., Delcher A.L., Utterback T.R., Weidman J.F., Khouri H.M.,
RA Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,
RA Frazer C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RT J. Bacteriol. 184:5479-5490 (2002).
DR EMBL; AB000516; AAK44530.1; -; Genomic DNA.
DR TIGR; MT0306; -;
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR000379; Ser_estrs.
DR Hydrolase; Hypochemical protein.

```



SEQ	SEQUENCE	400 AA;	44011 MM;	D6FD9510E84B8CB8 CRC64;
QY	Query Match	13.4%;	Score 293.5;	DB 2; Length 400;
Db	Best Local Similarity	30.5%;	Pred. No. 6.5e-14;	
Matches	97; Conservative	42;	Mismatches 122;	Indels 57; Gaps 14;
QY	95 AGLEYGNRRMLEKPSGFFAQP-----PLTEYAVAKVKDORRRSFYRIFPDGCF	143		
Db	84 SNLNFYAEIAAAAHDAKSPFAPTELPKYSRPASPLTEVAVAGTVD-----NIFAAGCF	137		
QY	144 TP-HGEGPQSGQSWLSYTNANRRYALLHHPH-PRPMLVCVGHSTEGRA.PLDLAVRANKL	201		
Db	138 RAINP-TYRQGSWSLTANNIVAAHWRHRDPRPTLLCYIHGF-MGSSYLLNGLEFSLPW	194		
QY	202 HDELGLNIVMPVLPNMGPRGQGLPKGAVFPG-----EDVLDDVHGAQAQVDIRLLSMI	256		
Db	195 YRSQSYDVLLATYLLPFGQRAE---KFSPPSGGFGYTSGLSGFAEMAGAVYDFRSIVLYL	251		
QY	257 RSQEEBSLIGLNGLSLGGYIASLVASLEBGLACAILGVPA-----DLIELL	303		
Db	252 RHIGVDR-IALTGISLGGYTSALLSVESRLEAVIPNCFVVPVPAKLPDEWFPXNLTVLK-	309		
QY	304 GRHGLRHKDPRRHVVKMAEPIGRMI-SPLSLTLPVMPGPGFYAGIADRLVHPREQVT	361		
Db	310 ---GLRLTN-----ISRDELTAGLVAGHPLNYRLPLDPRRMIIITGADRAAPEHAAT	360		
QY	362 RLMEHWGKPEIYVPGGH	379		
Db	361 -LWKQMDRCALHWPFGSH	377		
RESULT 7				
ID	053697 MYCTU	PRELIMINARY;	PRT;	400 AA.
AC	053697-			
DT	01-UN-1998 (TREMBlrel. 06, Created)			
DT	01-UN-1998 (TREMBlrel. 06, Last sequence update)			
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)			
DE	Hypothetical protein.			
GN	Ordered locus names=Rv0293c;			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacteriineae; Mycobacteriaceae; Mycobacterium;			
CC	Mycobacterium tuberculosis complex.			
OX	NCBI_TaxId=1773;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE.			
RC	STRAIN=H37RV;			
RC	MEDLINE=98295987; PubMed=9634220; DOI=10.1038/31159;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Church C.M.,			
RA	Harris D.E., Gordon S.V., Bismjelter K., Gas S., Barry C.E. III,			
RA	Tekela F., Badcock K., Basham D., Brown D., Chillingworth T.,			
RA	Connor R., Davies R.M., Devlin K., Fellwell T., Gentles S., Hamlin N.,			
RA	Holroyd S., Hornbly T., Jagels K., Krogh A., McLean J., Moutle S.,			
RA	Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,			
RA	Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Suleston J.E., Taylor K., Whitehead S., Barrall B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544 (1998).			
DR	EMBL; BX842573; CAA17368.1; -; Genomic_DNA.			
DR	PIR; C70837; C70837.			
DR	Tuberculist; Rv0293c; -			
DR	GO: GO:0016787; F:hydrolase activity; IEA.			
KW	InterPro; IPR000379; Set:sestr.			
KW	Complete proteome; Hydrolase; Hypothetical protein.			
SO	SEQUENCE 400 AA; 43971 MM; FBBBC136ADB4ABC CRC64;			
Query Match	13.4%;	Score 293.5;	DB 2; Length 400;	
Best Local Similarity	30.5%;	Pred. No. 6.5e-14;		
Matches	97; Conservative	42;	Mismatches 122;	Indels 57; Gaps 14;
95	AGLEYGNRRMLEKPSGFFAQP-----PLTEYAVAKVKDORRRSFYRIFPDGCF	143		
84	SNLNFYAEIAAAAHDAKSPFAPTELPKYSRPASPLTEVAVAGTVD-----NIFAAGCF	137		
144	TP-HGEGPQSGQSWLSYTNANRRYALLHHPH-PRPMLVCVGHSTEGRA.PLDLAVRANKL	201		
138	RAINP-TYRQGSWSLTANNIVAAHWRHRDPRPTLLCYIHGF-MGSSYLLNGLEFSLPW	194		
202	HDELGLNIVMPVLPNMGPRGQGLPKGAVFPG-----EDVLDDVHGAQAQVDIRLLSMI	256		
195	YRSQSYDVLLATYLLPFGQRAE---KFSPPSGGFGYTSGLSGFAEMAGAVYDFRSIVLYL	251		
257	RSQEEBSLIGLNGLSLGGYIASLVASLEBGLACAILGVPA-----DLIELL	303		
252	RHIGVDR-IALTGISLGGYTSALLSVESRLEAVIPNCFVVPVPAKLPDEWFPXNLTVLK-	309		
304	GRHGLRHKDPRRHVVKMAEPIGRMI-SPLSLTLPVMPGPGFYAGIADRLVHPREQVT	361		
310	---GLRLTN-----ISRDELTAGLVAGHPLNYRLPLDPRRMIIITGADRAAPEHAAT	360		
362	RLMEHWGKPEIYVPGGH	379		
361	-LWKQMDRCALHWPFGSH	377		
053697	MYCTU	PRELIMINARY;	PRT;	400 AA.
053697-				
01-UN-1998	(TREMBlrel. 06, Created)			
01-UN-1998	(TREMBlrel. 06, Last sequence update)			
01-OCT-2003	(TREMBlrel. 25, Last annotation update)			
Hypothetical	protein.			
Ordered locus	names=Rv0293c;			
Mycobacterium	tuberculosis.			
Bacteria;	Actinobacteria; Actinobacteridae; Actinomycetales;			
Corynebacteriineae;	Mycobacteriaceae; Mycobacterium;			
Mycobacterium	tuberculosis complex.			
NCBI_TaxId=1773;				
[1]				
NUCLEOTIDE	SEQUENCE.			
STRAIN=H37RV;				
MEDLINE=98295987				

Db	84	SNLNFYALAAAHDAKSPAPTELPKVTSRPASBLTWMVARGTVD-----NIAFASGF	137
Qy	144	TP-HFGEESGSGRMVLSYTNANREYALLLHPPE-PRWMLCVGHGTEMGRAPLDLAVPRAKTL	201
Db	138	RAINP--TMRQRMSALTANNIVHQAQWHRHROGPRFTLCVHGF-MGSSYLLNGLFFSLP	194
Qy	202	HDELGLNITVMPVLPNMGPRGQGLPKGAVFPG-----EDVLDDVHGTAQAVMDIRLLSMI	256
Db	195	YRSGYDVLATLTPFHGGRAB---KSPFSGHGYFTSGLSGFAEAMQAAYVDFRSIVDYL	251
Qy	257	RSQEBESLIGLNGLSLGGYIASLVASLEEGALCALGVPVA-----DLIELL	303
Db	252	RHIGVDR-IALTGLISLGGYTSALIASVESRLEAVIPNCPVMPAKLFDWEPPANKLVKL	309
Qy	304	GRHGGLRKHDPRRHVTKMAEPIGRMI--SPSLFLPYMPGPGFIIYAGIADRLVHPREQYT	361
Db	310	----GLRLTN-----ISRDELIALGALYHGPLNRYPLRLKDRRMITTGIDRNAPREHAYT	360
Qy	362	RLMEHMGKPEIWMYPGGH	379
Db	361	LMKQWDRCALHMPGSH	377

RESULT 8

ID	Q7U2C3	MYCBO	PRELIMINARY;	PRT;	400	AA.
AC	Q7U2C3					
DT	01-OCT-2003	(TREMBLrel. 25	Created)			
DT	01-OCT-2003	(TREMBLrel. 25,	Last sequence update)			
DT	01-MAR-2004	(TREMBLrel. 26,	Last annotation update)			
DE	Hypothetical protein MD0301c.					
GN	OrderedlocusNames=MD0301c;					
OS	Mycobacterium bovis.					
OC	Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;					
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium;					
OC	Mycobacterium tuberculosis complex.					
OX	NCBI_TaxID=1765;					
RN	[1]					
RP	NUCLEOTIDE SEQUENCE.					
RC	STRAIN=AF2122/97.					
RX	MEDLINE=22709107; PubMed=127889972; DOI=10.1073/pnas.1130426100;					
RA	Garnier T., Sigheleier K., Camus J.-C., Medina N., Mansoor H.,					
RA	Pryor M., Dutfoy S., Grondin S., Lacroix C., Monsempet C., Simon S.,					
RA	Harrie B., Actin R., Doggett J., Mayers R., Keating L., Wheeler P.R.,					
RA	Parthill J., Barrett B.G., Cole S.T., Gordon S.V., Hewinson R.G.;					
RT	"The complete genome sequence of Mycobacterium bovis."					
RL	Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).					
DR	EMBL; BX248335; CAD93165.1; -; Genomic DNA.					
DR	GO; GO:0016787; F:hydrolase activity; IEA.					
DR	InterPro; IPR000379; Ser estase.					
KX	Complete proteome; Hydrolase; Hypothetical protein.					
SO	SEQUENCE	400	AA; 43971	MM; FBBBC136A6DFABC	CR	C64;

Query Match 13.4%; Score 293.5; DB 2; Length 400;  
 Best Local Similarity 30.5%; Pred. No. 6.5e-14;  
 Matches 97; Conservative 42; Mismatches 122; Indels 57; Gaps 14;

AC Q1-2003 (Timblere1\_24, Created)  
 DT 01-JUN-2003 (Timblere1\_24, last sequence update)  
 DT 01-JUN-2003 (Timblere1\_24, last annotation update)  
 DE Hypothetical protein.  
 OS *Mycoccus xanthus*.  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;  
 CV Cytophactinae; Myxococcaceae; Myxococcus

OK NCBI\_TaxID=34;  
 RN NUCLEOTIDE SEQUENCE.  
 RA Hartzell P.L., Youderian P.A.;  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY044651; AAO22902.1; -; Genomic\_DNA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 326 AA; 35332 MW; C82D125BA41518DB CRC64;

Query Match 7.2%; Score 158; DB 2; Length 326;  
 Best Local Similarity 24.6%; Pred. No. 0.0011;  
 Matches 88; Conservative 36; Mismatches 138; Indels 96; Gaps 17;

QY 68 EIVLLGLKARRPVSNHRVF-----ERISQEVAAAGLEFYGN-----RRWLEK 108  
 DB 8 DVLPLGLSRR-----RLFSQGGDEQPLEEVAAAPFGQRPPIAPEMSAVRLQGLQV 62  
 QY 109 PEGFPAQPPPLTEVAVRKTKYKRRRSFYRLFPDSGFPFHGEGSQMWLS----- 157  
 DB 63 RDGTFPSPLARLDAAR-----TAH-----VWLSAGQPSRGACV 98  
 QY 158 YANNREYALTLRHPERPMLVCHGTEMGRAPLDLAVRAWKHDELGLNTVPLPMH 217  
 DB 99 VLAASREBSGFLRERYAP-----IAREGIDLF-----LENPYGAK 136  
 QY 218 GPRGQGLPKGAVFPGEDVDVHGTAAVMDIRLLSWIRSOEBSLIGLGLSLGYIA 277  
 DB 137 RVVGQ---KGAL--RTVSDHVLMLGVMDARALLAMRS--EGHARLGVAGYSGWYMA 190  
 QY 278 SLIVAS--LEEGALCALIGV--PYADLIE--LGRHCGLRKRD--PRRTTVKMAEPICMISP 331  
 DB 191 ALTAAVVPEPLVAVLAAGASPPVPTQGLSWSIAFALLDGRRAEARSRLGIFDL 250  
 QY 332 LSLTLPVP--MPGRFYAGIADRLVPRBOVTRLMHWKKEPIVWYPGSH--TGFFOSR 386  
 DB 251 ANLAPPPPRQPEAAVTVACRRDGFVGETTLALHAWMRSELRWADGHTALPTFR 308

## RESULT 12

09C7C5 ARATH PRELIMINARY; PRT; 375 AA.

AC 09C7C5;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein T21B14.3.  
 GN Name=T21B14.3;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 OC eurosida II; Brassicales; Brassicaceae; Arabidopsis.  
 OK NCBI\_TaxID=3702;

RP NUCLEOTIDE SEQUENCE.  
 RA Lin X., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,  
 RA Maiti R., Roming C.M., Koo H., Fujii C.Y., Uteback T.R.,  
 RA Barstead M.B., Bowman C.L., White O., Niernan W.C., Fraser C.M.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC069473; ANG51056.1; -; Genomic\_DNA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 375 AA; 42136 MW; A36E7F425875B6A2 CRC64;

Query Match 7.0%; Score 153.5; DB 2; Length 375;  
 Best Local Similarity 28.0%; Pred. No. 0.0028;  
 Matches 97; Conservative 40; Mismatches 131; Indels 79; Gaps 21;

QY 82 NHRVERISQEVAAAGLEFYGNRRWLEKPSGFFAQPPLTEVAVRKVKDRRSFYR--IFF 139  
 DB 39 NIELLERVQRLFP--LEVOG--QNW-----PPLVPRVWRVTMETKTATLREGVF-- 85  
 QY 140 DSGFTPHPE-----PGSQRWLSYTNANREYALTLRHPERPMLVCHGTEMGRAPL--DL 193  
 DB 86 ---QTPCADELTAALPPEBSR---TA--RVAMLVPRKVPQKMACVHLAVVARACLCDL 136

QY 194 AVF-----RAWKHDEL--GLNTVPLV--PMHGPRGQGLPKGA--VPEGVDLDDV 239  
 DB 137 NLFIPLPDGHTYDRLRLGGLVVKQNIATWVLESFFYQRPFLQCGRLLCVSDLL--L 194  
 QY 240 HGTAAVMDIRLLSWIRSOEBSLIGLGLSLGYIASLVASLEGLACALIGVPAVL 299  
 DB 195 LGRA--TIESRSLHWLDTBEGFGKMGVCGLSMGVHASMVGSLLH-----TPVALT 245  
 QY 300 IELLGRH-----C-----GLRHK--DPRRHVKAEPICRMISPLST-----P 336  
 DB 246 -PFLSPHAAVAFCEGLIKGTAWALREELAAQKITWTLDEVERMRNVSLTDVTRFP 304  
 QY 337 LVMPGFRFYAGIADRLVPRBOVTRLMHWKKEPIVWYPGSHGTF 383  
 DB 305 IPKNDAVIFVAATDDGYIPKHSVLELOKAMPGESEVRWVTGGHVSSF 351

## RESULT 13

09LH51 ARATH PRELIMINARY; PRT; 335 AA.

AC 09LH51;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Similarity to unknown protein.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 OC rosids; eurosida II; Brassicales; Brassicaceae; Arabidopsis.  
 OK NCBI\_TaxID=3702;

RN NUCLEOTIDE SEQUENCE.  
 RP MEDLINE=20363099; PubMed=10907853;  
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;  
 RA "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
 RT Sequence features of the 4,251,695 bp regions covered by 90 pl. TAC  
 RT and BAC clones.";  
 RL DNA Res. 7:217-221 (2000).  
 RN (2)

RP NUCLEOTIDE SEQUENCE.  
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP00063; BAB01965.1; -; Genomic\_DNA.  
 SQ SEQUENCE 335 AA; 37450 MW; D8AD777D6F1C27A CRC64;

Query Match 6.7%; Score 148; DB 2; Length 335;  
 Best Local Similarity 27.7%; Pred. No. 0.0064;  
 Matches 93; Conservative 38; Mismatches 133; Indels 72; Gaps 20;

QY 82 NHRVERISQEVAAAGLEFYGNRRWLEKPSGFFAQPPLTEVAVRKVKDRRSFYR--IFF 139  
 DB 14 NIELLERVQRLFP--LEVOG--QNW-----PPLVPRVWRVTMETKTATLREGVF-- 60  
 QY 140 DSGFTPHPE-----PGSQRWLSYTNANREYALTLRHPERPMLVCHGTEMGRAPL 194  
 DB 61 ---QTPCADELTAALPPEBSR---TA--RVAMLVPRKVPQKMACVHAGTGHYD-- 109  
 QY 195 VFRANKHDEL--GLNTVPLV--PMHGPRGQGLPKGA--VPEGVDLDDVHGTAAVMDIR 250  
 DB 110 --RRLRGGLVVKQNIATWVLESFFYQRPFLQCGRLLCVSDLL--LGRA--TIESR 164  
 QY 251 RLWSIRSOEBSLIGLGLSLGYIASLVASLEGLACALIGVPAVLIELGRH----- 306  
 DB 165 SLIHLDTBEGFGKMGVCGLSMGVHASMVGSLLH-----TPVALT--PFLSPHAAV 215  
 QY 307 --C-----GLRHK--DPRRHVKAEPICRMISPLST-----PLVMPGFRFYA 347  
 DB 216 AFCEGILKYGTAWALREELAAQKITWTLDEVERMRNVSLTDVTRPPKRPDAVIFV 275  
 QY 348 GIADRLVPRBOVTRLMHWKKEPIVWYPGSHGTF 383  
 DB 276 AATDGYIPKHSVLELOKAMPGESEVRWVTGGHVSSF 311

## RESULT 14

O9C7D8 ARATH

ID O9C7D8 ARATH PRELIMINARY; PRT: 360 AA.

AC O9C7D8; 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein P28J15.20.

GN Name=P28J15.20;

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI\_Taxid=3702;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,

RA Maiti R., Romling C.M., Koo H., Fujii C.Y., Utecherback T.R.,

RA Barnstead M.E., Bowman C.L., White O., Niernan W.C., Fraser C.M.;

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC069472; AAC51070.1; -; Genomic\_DNA.

KW Hypothetical protein.

SQ SEQUENCE 360 AA; 40396 MW; EA53832CD3C404C CRC64;

Query Match 6.7%; Score 148; DB 2; Length 360;

Best Local Similarity 27.7%; Pred. No. 0.0069;

Matches 93; Conservative 38; Mismatches 133; Indels 72; Gaps 20;

QY 82 NHRVERISQEVAAAGLEFYGNRRMLEKPSGFPAQPPPLETVAVRKVKDRRSFYR--IFP 139  
 DB 39 NLELLERMVQRLFP--LEVOG-QNW-----PPPLVRPVWRTWETKTATLREGVF- 85  
 QY 140 DSGFTPHPE-----PGSORMLSTYANNREYALLRHPEPRPLVCVHGTEGRAPLDLA 194  
 DB 86 ---QTPCADELTAALPPESR---TA--RVAMLVKRVPPQKACVHLAAGTGDHTYD-- 134  
 QY 195 VFRANKLHDEL--GLNIIVPVL--PMHGPRGQGLPKGA--VFPGEDVLDDVHGTAQAVMDIR 250  
 DB 135 --RLRLGSPLVKQNIATVLESFPGQRRPFLQCGARLLCVSDLL--LLGRA-TIEESR 189  
 QY 251 RLLSWIRSQEESLLGLNGLSLGGYIASLVASLEGACALLGVPAADIELLGRH---- 306  
 DB 190 SLIHWLDTBEGFGKMGVCGLSMGVHVASVGLHP-----TPVATL-PFLSPHSADV 240  
 QY 307 ---C-----GLRHK-DPRRHTVKAAPRIGRMISPLSLT-----PLVMPGRFIYA 347  
 DB 241 AECBGLIKKGTAWMALREBLAAQKITMTLDEVEREMRNVLSTDVTRPPIKRPDAVIFV 300  
 QY 348 GIADRLVHPRBOVTRLMHMGKPEIWMYPGGHTGFF 383  
 DB 301 AATDGYIPKHSVLELQKAMPSESVRWVTGCHVSSF 336

## RESULT 15

O8RWA3 ARATH

ID O8RWA3 ARATH PRELIMINARY; PRT: 363 AA.

AC O8RWA3; 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)

DE Hypothetical protein At3g12150 (At3g12156).

GN Name=At3g12150;

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI\_Taxid=3702;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,

RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,

RA Cheuk R., Chung M.K., Hayaishizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

RN [2]

RP NUCLEOTIDE SEQUENCE.

RA Kim C.J., Chen H., Cheuk R., Shinn P., Bowser L., Carninci P.,

RA Dale J.M., Hayaishizaki Y., Ishida J., Jones T., Kamiya A.,

RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,

RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,

RA Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,

RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;

RA Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.

RL EMBL; AY093234; AAM13233.1; -; mRNA.

DR EMBL; BT008851; AAP68290.1; -; mRNA.

KW Hypothetical protein.

SQ SEQUENCE 363 AA; 40784 MW; 6A86B799868E1748 CRC64;

Query Match 6.7%; Score 148; DB 2; Length 363;

Best Local Similarity 27.7%; Pred. No. 0.007;

Matches 93; Conservative 38; Mismatches 133; Indels 72; Gaps 20;

QY 82 NHRVERISQEVAAAGLEFYGNRRMLEKPSGFPAQPPPLETVAVRKVKDRRSFYR--IFP 139  
 DB 42 NLELLERMVQRLFP--LEVOG-QNW-----PPPLVRPVWRTWETKTATLREGVF- 88  
 QY 140 DSGFTPHPE-----PGSORMLSTYANNREYALLRHPEPRPLVCVHGTEGRAPLDLA 194  
 DB 89 ---QTPCADELTAALPPESR---TA--RVAMLVKRVPPQKACVHLAAGTGDHTYD-- 137  
 QY 195 VFRANKLHDEL--GLNIIVPVL--PMHGPRGQGLPKGA--VFPGEDVLDDVHGTAQAVMDIR 250  
 DB 136 --RLRLGSPLVKQNIATVLESFPGQRRPFLQCGARLLCVSDLL--LLGRA-TIEESR 192  
 QY 251 RLLSWIRSQEESLLGLNGLSLGGYIASLVASLEGACALLGVPAADIELLGRH---- 306  
 DB 193 SLIHWLDTBEGFGKMGVCGLSMGVHVASVGLHP-----TPVATL-PFLSPHSADV 243  
 QY 307 ---C-----GLRHK-DPRRHTVKAAPRIGRMISPLSLT-----PLVMPGRFIYA 347  
 DB 244 AECBGLIKKGTAWMALREBLAAQKITMTLDEVEREMRNVLSTDVTRPPIKRPDAVIFV 303  
 QY 348 GIADRLVHPRBOVTRLMHMGKPEIWMYPGGHTGFF 383  
 DB 304 AATDGYIPKHSVLELQKAMPSESVRWVTGCHVSSF 339

Search completed: March 23, 2006, 05:24:31  
 Job time : 108.63 secs

This Page Blank (uspto)



PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 7241  
LENGTH: 889  
TYPE: PRT  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-7241

Query Match 5.4%; Score 117.5; DB 2; Length 889;  
Best Local Similarity 22.8%; Pred. No. 0.0034;  
Matches 100; Conservative 59; Mismatches 158; Indels 121; Gaps 26;

QY 20 PVLGAMG-----PFGHTGLVVAQWRD-YIGQOPDKLPAPRTIALAQAQFDEIVLL 72  
DB 240 PVITALLNNAFGFIKIVIRGQL---SPADVFPDDPSKKP-----REAGSFIDQVGP 289  
QY 73 GL-----KARRVSNHRVPERISQEVAAGLEFYGRNRMLEKSGFAOPPLTEVAV 124  
DB 290 GLYLLEAGMGKTEBAALYAAYOMLVQERATGI-----YFALPTQLTSS--- 332  
QY 125 RKVKDRRSFYRIFPDGFTPHGPBGQRMWLTANNREYALLRHDEPRMVLVCVHGT 184  
DB 333 NKIDRFNAFLHQIV-SHETP-----QHSLL---HSGAMLM---DT 367  
QY 185 EMGRAPLDLAVFRWKHDELGL-----NIVMPVLPW-HG-PGQGLPKGAVFPG 232  
DB 368 EMGE---EGSPGGAWFNRKRGLLAPFVGTIDQALMVMVNHQGFVAYGLAGKV--- 421  
QY 233 EDVLDVH-----GT-AQAVWDIR-----LISWRSQ-EESLSGLGSLGCIYA 277  
DB 422 --ILDEVHTYDLVTGTIINALVEFLRQIDCVIILSATLSQTRDALLQOSTSEAYPLI 479  
QY 278 SLVSLSEGLACATIGVPVADILIELGHRGRLHKDPRRHVTKMAEPIGRMISPLST-- 335  
DB 480 TAASBAEREGVIGVGVTEHTTYILHSC-RKQEPAREELRAELGQVWMENTIA 537  
QY 336 ---PLVMPGRFIYAGIADRLVHR--EQVTRLMHWKPEIIVWPGHGTGFQSPRV 388  
DB 538 EAQGYLDLASRAVAGETGLHRSFTPOHNRHREQW---VALY--GRAGMPQRKQC 591  
QY 389 RRFV--QAALFQSGILDA 404  
DB 592 GRIVGTQVLEOSLIDA 609

## RESULT 3

US-09-252-991A-20236  
Sequence 20236, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 20236  
LENGTH: 409  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20236

Query Match 5.1%; Score 112.5; DB 2; Length 409;  
Best Local Similarity 24.7%; Pred. No. 0.0035;  
Matches 94; Conservative 33; Mismatches 111; Indels 143; Gaps 23;  
QY 105 WLEKPSGFFAOPPLTEVAVRKVKDRRSFYRIFPDG---FTPHGPBGQRMWLT-- 158

DB 66 WDEYATGVIRRHPR-----ADNRK---RLPALGPRTATLEGHPG-GGSVPMEGYAT 114  
QY 159 -----TANREYALLRHPRP-----RPMVLVCV----- 181  
DB 115 GVIRRYPAATDNKRKLEFALQAHVPCRKATGMPFCMAITATGAIKCLPAVLHRHRAQV 174  
QY 182 ---HGTMGRAPLDLAVFRWKHDELGLIYN--PVL-----PMGPR-----GQ 222  
DB 175 DTDHHRVVVG-GPVRLGVQAVLLPRGLELDVVVQHGLGVGVDPVPLRAAYLARQAGGE 233  
QY 223 GLPKGAVFPGSDV-----LDVH-----GTQAVWDIR--RLSWSIR 257  
DB 234 GVVAVVAGLPGSHVIALSCARLTQHQVAGAADVHRVFLGIGIEVADDQATRLIGARR 293  
QY 258 SQEESLSGLGSLGCIYASLVASLEGACATIGVPVADILIELGRHGLR--HKDPR 315  
DB 294 IARQPVHGLGAGARG-----QVAVLVAVAGRVA---VABRALRLQVVRHDCQ 339  
QY 316 RH-TVKMAEPIGRMISPLSTPLVMPGRFTYAGIADRLVHPRQVTRLMWH---WGP 370  
DB 340 AFAAIDLEGLGQ-----GRTV-AGVVELRVH-----RTVEHRRGSIGAD 378  
QY 371 EIVWYPGHGTGF--QSRPVR 389  
DB 379 LVAALDPHGNFVAALRPVR 399

## RESULT 4

US-09-252-991A-26849  
Sequence 26849, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 26849  
LENGTH: 1127  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26849

Query Match 5.0%; Score 109.5; DB 2; Length 1127;  
Best Local Similarity 23.5%; Pred. No. 0.036;  
Matches 84; Conservative 36; Mismatches 100; Indels 137; Gaps 18;

QY 123 AVRKVKDRRSFYRIFPDGFTPHGPBGQRMWLTANNREYALLRHPRP--WLV 179  
DB 27 AARSIPSR-----PGRPCNR---CWTTRSVPPLPFGSAFASAWW-- 65  
QY 180 CVHGTMGRAPLDLAVFRWKH--DEGLNIVMPVLPMHGPBGGLPKGAV-----FPG 232  
DB 66 ---SRSGRCPRDSCFTSRLFHYDQERLEIRLGIFVANEHGGGIGQEAVALRAGAARV 121  
QY 233 EDVLDVHGTAAQAVWDIRRLRLSWIRSQEESSL---IGNLGSLGCIYASLVASLEGSLA 288  
DB 122 DRDADVHGLA-----VHLQRADLIGHNRRLDVAAVGGDL-DLVA---GID 164  
QY 289 CATIGVPVADILIELGHRGRLHKDPRRHVTKMAEPIGRMISPLSTPLV--PMGR 343  
DB 165 AQLGQRLADLDELGL-----LGDRLQPAVLGVVLEGEVPGGR 204  
QY 344 FI-----YAGIADRL-----VHPREOVTRL-----WEHMGKPEI 372  
DB 205 RVGELVGRABQFHVVLHPRCARVADRLAVALVAGVHPRRLERLVVLGERRAFGLVDBGQ 264

Qy 373 WYPCG-----HTGFQSRPVRRFQAL-----ESGLDAPRTORD 410  
Db 265 ASHPGVADDERVHPGF-----RRTVGLVGVHGAAPGLAVPRHQAALLRRVRLAAD 315

## RESULT 5

US-09-976-674-43  
Sequence 43, Application US/09976674  
Patent No. 6844180  
GENERAL INFORMATION:  
APPLICANT: Q1, Steve  
APPLICANT: Akimsanya, Karen  
APPLICANT: Riviere, Pierre  
APPLICANT: Junien, Jean-Louis  
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPLV  
FILE REFERENCE: 70669  
CURRENT APPLICATION NUMBER: US/09/976,674  
CURRENT FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: US 60/240,117  
PRIOR FILING DATE: 2000-10-12  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 43  
LENGTH: 691  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-976-674-43

Query Match 4.9%; Score 106.5; DB 2; Length 691;  
Best Local Similarity 18.2%; Pred. No. 0.035;  
Matches 63; Conservative 63; Mismatches 106; Indels 115; Gaps 14;

Qy 9 THERSAFRLSPVYLSGANGPFMTGLVYAQSRDYLGQOPDLPIARPTIALAAQAFRDE 68  
Db 288 SRDGSKFMTVYVVKQSGRSEFHHIMFLIQSKSEOI-----TVRHLTSGNNEV 335  
Qy 69 IYVLGLKRRPVSNNRVRFERISOEVAAGLEFYGNRRMLEKSGFPAQPPRLTEVAVRKVK 128  
Db 336 IKIL-----AYDETQKISASTEGALNRQC-----ISCNFMK 367  
Qy 129 DRRSFYRIFPDSGFTPHGEPBSQRMWLSY-----TANNEVALL----- 168  
Db 368 EQ-----CTYFASISPM-----NQHFLFCGPRPVVSLHSTONPAKYFLLENSMLK 417  
Qy 169 ---LHPRPRLVVCVHGTGWRAPLDLAVFRAMKLHDELGLNIWVPLPMHGPRGQLP 225  
Db 418 EAILKKIKGPKRIKILHIDY-ELPLQLSLPRDFMRNQYALLIMD-----EPPGQLVT 472  
Qy 226 -----KGAVPRGVDVDDVH-----GTAQAVMDIRL-----L 253  
Db 473 DKFHIDWDSVLIDMNVIARFDPGRSGSGFQGLKILQEHRRIGSVAVKQDITAVXFLKL 532  
Qy 254 SWIRSOEBSLIGLNGLSIGYIASIVASLEBGL-ACAILGVAVDL 299  
Db 533 PYIDSKR-----LSIFGKGYGTYIASMILKSDKLRKCGSVAVPIIDL 575

## RESULT 6

US-08-764-100-27  
Sequence 27, Application US/08764100  
Patent No. 5773700  
GENERAL INFORMATION:  
APPLICANT: van Grinsven J., Martinus O.  
APPLICANT: De Haan, Petrus T.  
APPLICANT: Gielen L., Johannes J.  
APPLICANT: Peters, Dirk  
APPLICANT: Goldbach, Robert W.  
TITLE OF INVENTION: Improvements in or Relating to Organic  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sandoz Agro, Inc

STREET: 975 California Avenue  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/764,100  
FILING DATE: 06-DEC-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,064  
FILING DATE:

APPLICATION NUMBER: US 08/032,235  
FILING DATE: 17-MAR-1993  
APPLICATION NUMBER: GB 9206016.9  
FILING DATE: 19-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 577370019, Allen E.  
REGISTRATION NUMBER: 34,490  
REFERENCE/DOCKET NUMBER: 137-1061  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 354-3592  
TELEFAX: (415) 857-1125

INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3218 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-764-100-27

Query Match 4.9%; Score 106.5; DB 1; Length 3218;  
Best Local Similarity 22.4%; Pred. No. 0.039;  
Matches 81; Conservative 40; Mismatches 158; Indels 83; Gaps 14;

Qy 33 GLVYAQSRDYLGQOPDLPIAR--PTIALAAQAFRDEIYVLGLKRRPVSNNRVRERI- 89  
Db 2069 GLYALATHRCYSGLYPHCYSARGASNVALTYRASPMETASPLGNSRPHARGILETYR 2128  
Qy 90 SOEVAAGLEFYGNRRW-----LEKPGFPAQPPRLTEVAVRKVKD 129  
Db 2129 SERVALILIELVSSERTIRILVSSERGLVALCYSLSESGLYPH-----EVAL----- 2178  
Qy 130 RRRSFYRIFPDSGFTPHGEPBSQRMWLSY-----TANNEVALL-----LHPRPRLVVCVHGTGWR 187  
Db 2179 ---GLYIALALYSCSPHETHTVALSERGLG-----LVALPRSRG 2215  
Qy 188 RAPDLAVFRAMKLHDELGLNIWVPLPMH-----GPRGQLPKGAVPRG-----EDVIDD 238  
Db 2273 LYGLNLIELLALHNSIGYPSPASPEERHISVALTYRALAGYASVILRLAARGLEASNA 2332  
Qy 239 VHG-----TAQAVMDIRLLSWIRSOEBSLIGLNGLSIGYIAS-IVASLEBGLACAIL 292  
Db 2273 LYGLNLIELLALHNSIGYPSPASPEERHISVALTYRALAGYASVILRLAARGLEASNA 2332  
Qy 293 GAVPAD-LIELGRHCGLRHKOPRRHTVYKMAEPIGRMISPLTFPLVMPGFRFIYAGIAD 351  
Db 2333 SNFSEERLYSMETPHEGLYHISPRQJLSESRHIS-----GLNGLYASPRRIIEPHSE 2386  
Qy 352 RL 353  
Db 2387 RL 2388

## RESULT 7

US-09-252-991A-17549  
Sequence 17549, Application US/09252991A



Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 17549  
; LENGTH: 563  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17549

Query Match 4.6%; Score 102; DB 2; Length 563;  
Best Local Similarity 23.7%; Pred. No. 0.078;  
Matches 87; Conservative 36; Mismatches 106; Indels 138; Gaps 22;

QY 20 PVLGANGCPFHNTGLYANQW-----RDYL-CGQPD-----KLPIAR-PTIALAAQAR- 66  
DB 114 PVLGSG-----IQPWPVDSRDPDLRADQGDAAAGRLGVALLPRLALREAGRR 163  
QY 67 --DEIVLLGLKARRPVSN-HRVFERISQEVAAGLEFYGNRRRLKESGFAPQPPPL----- 119  
DB 164 LDDRPRRGGRKRRKVRRAHL-----GRRIHPQRPAPRIPGPPVAPRT 208  
QY 120 ----TEVAARVKVKKRRRSFYRIFPDQSGFTPHPGEGSGRWLSYANNREYALLRHPEPR 175  
DB 209 HQPGRALRLQRLDRRAV-----RLGQP-----RPAIHLGQPEPG 245  
QY 176 PWLVCHNTGEMGRPLDAVFRANKHDELGLNIVPPLP--MGPRQGGPK-----GAV 229  
DB 246 GAVGACPRRGERRPARLD-----PAPRLRGLPQGGPRILAAAGAG 288  
QY 230 FPGSDVLDVHGTQAQVWDIRLLSWIRSOEESLIGLNGSLGQYIASLASEGLAC 289  
DB 289 RPA-----LHALAAL-----ARAAMLRACAG-----GAGG--AAKRLPAGARRAC 326  
QY 290 AILGVPVADLIELLGRHCGLRHKDPRRHTVKAEDIGMISPLSTPLVPMGRFTYAGI 349  
DB 327 AGSGILLVLL-----AKTH-----PPLRRRRRP-ALSPVVLPVPA--VAGGL 365  
QY 350 ADRLVHP 356  
DB 366 LPALERP 372

RESULT 8  
US-09-902-540-16812  
; Sequence 16812, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Steven C.  
; APPLICANT: Slater, Gregory J.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(115849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 16812  
; LENGTH: 981  
; TYPE: PRT  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-16812

Query Match 4.5%; Score 99.5; DB 2; Length 981;  
Best Local Similarity 21.4%; Pred. No. 0.34;  
Matches 66; Conservative 41; Mismatches 105; Indels 97; Gaps 14;

QY 107 EKPSGFAQAPPLP-EVAVKRVKKRRRSFYRIFPDQSGFTPHPGEGSGRWLSYANNREY 165  
DB 718 ELPS--VAKPEPTPNVQKQVPER-----FTSLVPRDFKPKK----- 757  
QY 166 ALLRHPEPPWLVCHNTGEMGRAPDLAVFRANKH-----DE--LGLNIVPVLPM 216  
DB 758 -----LPVIVEYG-----GPTTVVHKSAALMSQVADQGFVLVKIDGRGRL 803  
QY 217 HGRPGGLPKGAVPGEVDVHGTQAQVWDIRLLSWIRSOEESLIGLNGSLGQYI 276  
DB 804 RGAKTREVXGD-FSGVTLEDOVALQALAKEVEL-----DLRVGISGWSFGGYM 854  
QY 277 ASL-VASLEGLACAILGVPVADLIE-----LGRHCGLRHKDPR----- 316  
DB 855 SALAVLKRDPVFKAGVAGAPVWMLDYDTHTERYLGVPOQSPRAYEKSSLLTYAKODKP 914  
QY 317 -----HTVKAEPYGRMISPLSTPLVPMGRFTYAGIADRLVHPRE 358  
DB 915 MGKLLIHGTADDVVFHTLKLSDALFRAGKPHDLPLSGLTIM-----VPDELVTER- 968  
QY 359 QVTRLWEHW 367  
DB 969 QWERVMDHF 977

RESULT 9  
US-09-902-540-13877  
; Sequence 13877, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Steven C.  
; APPLICANT: Slater, Gregory J.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(115849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 13877  
; LENGTH: 748  
; TYPE: PRT  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-13877

Query Match 4.5%; Score 98.5; DB 2; Length 748;  
Best Local Similarity 21.7%; Pred. No. 0.29; Indels 137; Gaps 23;  
Matches 100; Conservative 52; Mismatches 171;

QY 40 WRDYLGOQPPKLPAPRTIALAAQAFRDEIVLLGLKAR----- 77  
DB 84 WKRYLGARPLPLP--DTLTFEGRAY-----TVGVWMESEFDYRGQLWVRELVPRLPS 136  
QY 78 RPVSNNRVFERISOEV--AAGLEFYGNRRRLKESGFAPQPPPLTVAVRAKYKDR----- 130  
DB 137 RTAHNWRVVRQLADGVLLQARVELTHIARELENOYG---QDRTMHDAIVAEPLQOESLVGR 193  
QY 131 -RBSFYRIFPDQSGFTPHPGEGSGRWLSYANNREYALLRHPEPPWLVCHNTGEMGR 189  
DB 194 VRSTLYLL-----AGAAFLLLVAGAVNTLLAKRAATRABELAIH-VALGAG 240  
QY 190 P-----LDL-----AVFRANKHDELGLNIVPVLPMHGR----- 220  
DB 241 PGALMRRLMESLUSLISLAGALGAVLAAW-----GVRAALLAEPHGLPRVDEVGVNATV 294  
QY 221 -----GGGL-----PKGAVFGEEDVLDDVHGTQAQVWDIRLLSWIRS 258

```
Db      295 LIFSLGLSLALLAGLGLTALRAARPPGAAALAGSGRTSGGSAART---RRL--VVG 349
Qy      259 QEEESLIGNGSLGVIASIVASLEBGLA---CALLGV---PVADIELLGRHGLRHK 312
Db      350 QALALVLLVGAALLGRSLMGLSLDPCGRTRETVAVLSLVLPAAK--EAGGRH---NV 403
Qy      313 DRRRTVK--MAEPIGRMISPLSLPLVMP--GRFIYAGIADRLVHPRE--QVTLMWE 365
Db      404 QLOEHLISRLAALPGVAVGAVSSPFLGSPAGDGFTIYLNRPDEVGNFEDFGRLAREPE 463
Qy      366 HNGKPEIYVYPGHNGFPGSRPRRVRQAALLESGLDLP 405
Db      464 RTGSAAE---YRVASEGYFALGIPLVGRGLFDARDTVDAF 500

RESULT 10
US-09-902-540-15329
; Sequence 15329, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 15329
; LENGTH: 1178
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-15329

Query Match      4.5%; Score 98.5; DB 2; Length 1178;
Best Local Similarity 23.7%; Pred. No. 0.59;
Matches 85; Conservative 36; Mismatches 109; Indels 129; Gaps 20;

Qy      17 LSPVLSGAMGPFMTGLVVAQSWRDYLGQDPKLPAPRTTALAQAARDEI----- 69
Db      785 VAPF--GAPMFT-----RDLASVSDALKER-----FQALGHELOAKLRD 825
Qy      70 -----VLGLKARRP---VSNHR-----VFERIS--OEVAAGLE-----F 99
Db      826 GGPVVTTLLELGGPRRPSRLMTAHRLLVVDGVSWRLFFQOLLAVEQLGGEGECKLPRT 885
Qy      100 YGNRRMLEKSGGFPAPQPLTE-----VAVRKVKDRRARSFYRIFPDGFTPHGSRGS-Q 153
Db      886 TSVKMLERLQG-VASPRRLREELAHMTAQGKQRPPLVER-----PGLVFTSLSE 936
Qy      154 RMLSTANNREYALLLRH--PEBRPMLVVCVHTEMGRAPDLAVFPAKMLHDELGINIWP 212
Db      937 RELTALASSESRLLINLEPE-----AYRCEIGEL--LATLAWTHTWGTGHPVLL 986
Qy      213 VLPWNGPRGQGLPKGAVPFGEDVLDVHGTAQAVMDIRRLSWINS----- 258
Db      987 DIEGHG-----REDVFEVDY-----DLSTRVGLTSLVPRLTTPRVRGS 1025
Qy      259 -----QEEESLIGNGSLGVIASIVASLEBGLACALISGPVADLIELGRHGLRHK 311
Db      1026 AATVRTVQQLRSVPRKGL--GIVGLRYISPDALRLARLARAPQA---EVLFRYGGQYH 1078

RESULT 11
US-09-602-777A-118
; Sequence 118, Application US/09602777A
; Patent No. 6831165
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
```

```
; APPLICANT: Kroege, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberbauer, Gregor
; TITLE OF INVENTION: Corynebacterium glutamicum GENES ENCODING PROTEINS
; FILE REFERENCE: BGI-128CP
; CURRENT APPLICATION NUMBER: US/09/602,777A
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: DE 19931636.8
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932126.4
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932127.2
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932129.9
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: DE 19932226.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932920.6
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932924.9
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932928.1
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932930.3
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932933.8
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932935.4
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932973.7
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19933002.6
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19933003.4
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19933005.0
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19933006.9
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19941378.9
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19941379.7
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19941390.8
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19941391.6
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19942088.2
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 442
; SEQ ID NO 118
; LENGTH: 668
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-602-777A-118

Query Match      4.5%; Score 98; DB 2; Length 668;
Best Local Similarity 20.8%; Pred. No. 0.27;
Matches 91; Conservative 52; Mismatches 157; Indels 138; Gaps 25;

Qy      41 RDVYGGQDPKLPAPRTTALAQAFA-----RDEIYGLKARRPVSNHR- 84
Db      289 KEFLGGKDFQPVFPFTSTSLGSLATTKNPLVLTLLNNVSTELVTVPL--NDPTTEHH 346
```

```
QY 85 --VERISOEVAAGLEFYGNRRLEKPSGFFAOPPLTEV---AVRKVORRRSFYRI 138
DB 347 IDLEPHVAHVAVATSPDODEIIVQAA--FTEAPTLRAELPGALEAVKAPLQF---- 400
QY 139 FDSGTFPHGEGSQRWLSYTANREYALLR---HPERPVLVCHGTEKRAPLDLAV 195
DB 401 -----ENAGQETRQHWAT-SADGKIPYFITGAEEBEPONTLVHAYG----- 441
QY 196 FRAWKLDHDELGINIWPVLPWNGP-RG-QGLPKGAVF-----PSADVLDVHGTA--- 243
DB 442 -----GEVSLTPSHSPTRGIAWLEKGYTFVEANRGGEPEPHSQATKLN 489
QY 244 -QAVW-DIRRLSWI--RSOEESLIGLNGSLGGYIAS-LVASDEGLACAILGVPAV 298
DB 490 RMKWEHRAVLADLVERGYATPEQIAIRGSGNGGLTSGALTQYPEAFGAIVGVPLAD 549
QY 299 LIEELGRHCGLRHKDPRRHTTK-----MAE-----PIGRMI-----SPLSLTPVLP--M 340
DB 550 ML-----RHYTWSAGASWMAEYGNPDPPEERAVIEQYSPQAVVGEKRIY 595
QY 341 PGRFIYAGIADRLVHPRQVTRLM-----EHMGKPEIWMVPGGHTS-----F 382
DB 596 PPALVTSTTRDVRPAH--ARLFPAQLIDAGQAVDYENTE-----GSHAGAADNKQTA 648
QY 383 FQSRPVRREFVQAALEQSG 400
DB 649 FVESLIYTWIEKTLDDQG 666
```

RESULT 12  
US-09-105-537-22  
Sequence 22, Application US/09105537A  
Patent No. 6265202

```
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/105,537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 22
LENGTH: 402
TYPE: PRT
ORGANISM: Streptomyces venezuelae
US-09-105-537-22
```

Query Match 4.4%; Score 97; DB 2; Length 402;

Best Local Similarity 22.8%; Pred. No. 0.16;

Matches 75; Conservative 35; Mismatches 139; Indels 80; Gaps 15;

```
QY 110 SGFFAOPPLTEVAVRKVDRRRSFYRIFFDSGTFPHGEGSQRWLSYTANREYALL 169
DB 6 TGAULTQPP--LGRITRAVADRE-----LGTILLETRGIHWI--HAANGDPYATVL 51
QY 170 RHPERPVLVCHGTEKRAPLDLAVFRWKLDH--LGINIWPV-LPHNGPRQGLPKG 227
DB 52 RQADDPY--PAYEVRARAGALSPTSQSWTADHAAASILCSTDGVSADGVVPVPOQ 109
QY 228 AVFPGEDVLDVHGTAVMDI-----RRLSWIRSOEESLIGLNGSL-----LGGYIA 277
DB 110 VLSYEGCGRLERQVLPAGADVPEGGQRAVVEGIRHETIEGLADPPSASVAFELLGGFVR 169
QY 278 SLVASLEGLACAILGVV-----ADLIELGRHCGLRHKDPRRHTTKMAEPI----- 325
DB 170 PAVTA-----AAAAVLGVPRADPADLLERLPLSDSLAPQSLRTVAADGALAEELTA 225
QY 326 -----GRMISPLSLTPVLPWNGRFTYAGIADRLVHPRQVTRLMENHGKPEIWMV 376
```

```
DB 226 LIAOSDSDPGALLSLNLTAVOLTGNAVLA-----LLANP-----EQW--RELCDRP 271
QY 377 GGHGTFQSRPVRREFVQAALEQSGLLDAP 405
DB 272 G-----LAAAABEETLRYDPP 287
```

RESULT 13  
US-09-105-537-4  
Sequence 4, Application US/09105537A  
Patent No. 6265202

```
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/105,537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 3782
TYPE: PRT
ORGANISM: Streptomyces venezuelae
US-09-105-537-4
```

Query Match 4.4%; Score 97; DB 2; Length 3782;

Best Local Similarity 22.8%; Pred. No. 5.3;

Matches 75; Conservative 35; Mismatches 139; Indels 80; Gaps 15;

```
QY 110 SGFFAOPPLTEVAVRKVDRRRSFYRIFFDSGTFPHGEGSQRWLSYTANREYALL 169
DB 6 TGAULTQPP--LGRITRAVADRE-----LGTILLETRGIHWI--HAANGDPYATVL 51
QY 170 RHPERPVLVCHGTEKRAPLDLAVFRWKLDH--LGINIWPV-LPHNGPRQGLPKG 227
DB 52 RQADDPY--PAYEVRARAGALSPTSQSWTADHAAASILCSTDGVSADGVVPVPOQ 109
QY 228 AVFPGEDVLDVHGTAVMDI-----RRLSWIRSOEESLIGLNGSL-----LGGYIA 277
DB 110 VLSYEGCGRLERQVLPAGADVPEGGQRAVVEGIRHETIEGLADPPSASVAFELLGGFVR 169
QY 278 SLVASLEGLACAILGVV-----ADLIELGRHCGLRHKDPRRHTTKMAEPI----- 325
DB 170 PAVTA-----AAAAVLGVPRADPADLLERLPLSDSLAPQSLRTVAADGALAEELTA 225
QY 326 -----GRMISPLSLTPVLPWNGRFTYAGIADRLVHPRQVTRLMENHGKPEIWMV 376
DB 226 LIAOSDSDPGALLSLNLTAVOLTGNAVLA-----LLANP-----EQW--RELCDRP 271
QY 377 GGHGTFQSRPVRREFVQAALEQSGLLDAP 405
DB 272 G-----LAAAABEETLRYDPP 287
```

RESULT 14  
US-09-902-540-10121

Sequence 10121, Application US/09902540

Patent No. 6833447

GENERAL INFORMATION:

APPLICANT: Goldman, Barry S.

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Wiegand, Roger C.

TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof

FILE REFERENCE: 38-10(115849)B

CURRENT APPLICATION NUMBER: US/09/902,540

PRIOR FILING DATE: 2001-07-10

PRIOR APPLICATION NUMBER: 60/217,883

NUMBER OF SEQ ID NOS: 16825

```

; SEQ ID NO 10121
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-10121

```

```

Query Match      4.4%; Score 96.5; DB 2; Length 596;
Best Local Similarity 23.7%; Pred. No. 0.33;
Matches 97; Conservative 38; Mismatches 140; Indels 135; Gaps 23;

```

```

QY 8 GTHERSAFRLSPVLSGANGPFWHTGLVYAQSWRDY---LGOQPDILPIAR----- 55
DB 169 GTHKS-----LGDARPSLIRRRRHGHLMRDRIIGRELDLGVLRKVGNGPQGF 220
QY 56 ---PTALAA---QAFDEIVLLGLKARRPVSNHRVFERISQEVAAGLEFYGNRRMLE- 107
DB 221 VPPPSIGMAAWLOTATRRPALLAKAE---CARGFQVSVR-----NRPVAA 268
QY 108 -KPSGFPAQ---PPLTEVAVRKVKDRRRSFYRIIPDSGFTPHGPEPSQRMVLYTANN 162
DB 269 PPDGQLLPERRVAPYFEEYAVPDARQAARFVHEFV-----EP----- 307
QY 163 REVALLR---HPEPRMLVCV---HGTENGRAPLDLA-----VPRAMKLDDELG 206
DB 308 -----LRDAVRAEPFPYVACVADGDGRGRLAEALAROPDGHAAHQRLSRALADFTROA 361
QY 207 LNIIV---MPVLPMHGRPGQLPKGAVFGEDEVLD-----DVHGTAAQVMDIRLLSMIR 257
DB 362 KTIIVEVHNGVLVYAG-----GDDILAFVTPMDALACQA-----LAVAFR 402
QY 258 SQEESLIGLN-----GLSLGGYIASLVASLEEGIAC--AIIIGVPVADLITELLGRHC 307
DB 403 ATLEALAGTRAEVPTLVGGLIGVHLESLGELLERGRAERAAXKODGRDALAVLVAKHA 462
QY 308 GLRHKDPRRHTVMA--EPIGRMISPLSL--TPLVPMGRFTY--AGIADR 352
DB 463 GRE---RLMTAPWATDPVERLTRDMALLSSDTQPLKKVHEVVALARR 508

```

```

RESULT 15
US-09-902-540-15882
; Sequence 15882, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 15882
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-15882

```

```

Query Match      4.4%; Score 96; DB 2; Length 416;
Best Local Similarity 24.1%; Pred. No. 0.21;
Matches 76; Conservative 35; Mismatches 115; Indels 90; Gaps 18;

```

```

QY 8 GTHERSAFRLSPVLSG---AMGP---FMHT-----GLVYA-----QSWRDYLGQO 47
DB 76 GEVDTSADARALPYVLSGCGVLRIGAFIRTDATAGTPVGLANVSLDQGRQILTL 135
QY 48 PDKLP--IARPT--ALAAQAFRDE-----IVLLGLKARRPVSNHRVFERISQEVAAGLEF 99
DB 136 ADALKGVABPAAASALPVQDTTIDEAERIRAILLAIST--AGNKGVLRRRGRTVRLVLR 192

```

```

QY 100 YG-----NRRW-LEKPSGFPAQPPPLTEVAVRKVKDRRRSFYRIIPDSGFTPHGEGSOR 154
DB 193 YNAETEOLEMAHEDPYTDMGEAPYDIDVI-----GHNSAYMRLEA-----GTPGQGR 240
QY 155 WLSYTAANNREVALLLRH-----PEPRPMLVCVHGT-----EMGRAPLDLA---VFRAM 199
DB 241 ---FTSPPLDVLFRIHRKRRSPAPBEGVVTFFHHPLMREGEMERAVLDLSFSGMCIRC 297
QY 200 KLHDELGLNIYMPVLPMHGPRGQLPKGAVFGEDEVLDVHGTAAQVMDIRLLSMIRSQ 259
DB 298 RPEDLLFPGLLPPLLELHTPPDQSI-----SLRGEIRYVTS 334
QY 260 BEESLIGLNGSLGGY 275
DB 335 RADGTV-LCGLSVSPY 349

```

Search completed: March 23, 2006, 05:30:19  
Job time : 29.268 sec

This Page Blank (usptc

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using SW model

Run on: March 23, 2006, 06:14:33 ; Search time 86.8513 Seconds  
(without alignments)  
1985.885 Million cell updates/sec

Title: US-10-617-038-29

Perfect score: 2195  
Sequence: 1 MASSASDGTHERSAPRLSP.....AAEGLDAPRTORDRSA 413

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA\_Main:\*  
1: /cgn2\_6/pcodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/pcodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/pcodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/pcodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
5: /cgn2\_6/pcodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
6: /cgn2\_6/pcodata/1/pubppaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2195	100.0	413	4 US-10-617-038-29	Sequence 29, Appl
2	146.5	6.7	366	4 US-10-425-115-286397	Sequence 286397,
3	146.5	6.7	369	4 US-10-425-114-39123	Sequence 39123, A
4	146.5	6.7	369	4 US-10-425-114-47261	Sequence 47261, A
5	130	5.9	303	4 US-10-424-599-196342	Sequence 196342,
6	121	5.5	338	6 US-11-097-143-27990	Sequence 27990, A
7	118.5	5.4	322	5 US-10-739-930-9602	Sequence 9602, Ap
8	112	5.1	574	4 US-10-282-122A-62026	Sequence 62026, A
9	111.5	5.1	311	4 US-10-437-963-183563	Sequence 183563,
10	110	5.0	481	4 US-10-424-599-227907	Sequence 227907,
11	107.5	4.9	1470	4 US-10-425-115-206590	Sequence 206590,
12	106.5	4.9	691	3 US-09-876-674-43	Sequence 43, Appl
13	106.5	4.9	691	5 US-10-982-512-43	Sequence 43, Appl
14	106	4.8	452	4 US-10-425-114-60109	Sequence 60109, A
15	104.5	4.8	742	4 US-10-369-493-15631	Sequence 15631, A
16	104.5	4.8	742	4 US-10-369-493-16006	Sequence 16006, A
17	104.5	4.8	742	4 US-10-369-493-16360	Sequence 16360, A
18	103	4.7	810	4 US-10-425-115-274900	Sequence 274900,
19	102.5	4.7	980	5 US-10-794-514A-342	Sequence 342, App
20	100.5	4.6	586	5 US-10-755-415-241	Sequence 241, App
21	100	4.6	740	4 US-10-437-963-115057	Sequence 115057,
22	99.5	4.5	424	4 US-10-425-114-64299	Sequence 64299, A
23	98.5	4.5	1415	4 US-10-425-115-206589	Sequence 206589,
24	98	4.5	296	4 US-10-425-115-224377	Sequence 224377,
25	98	4.5	668	5 US-10-721-922A-178	Sequence 178, App
26	98	4.5	706	3 US-09-738-626-3879	Sequence 3879, Ap
27	98	4.5	706	5 US-10-494-541-22	Sequence 22, Appl

28	97	4.4	402	3 US-09-861-289-22	Sequence 22, Appl
29	97	4.4	402	3 US-09-860-846-22	Sequence 22, Appl
30	97	4.4	402	3 US-09-988-384B-22	Sequence 22, Appl
31	97	4.4	402	3 US-09-836-821-22	Sequence 22, Appl
32	97	4.4	402	4 US-10-271-889-22	Sequence 22, Appl
33	97	4.4	402	4 US-10-398-605-22	Sequence 22, Appl
34	97	4.4	706	4 US-10-425-115-309560	Sequence 309560,
35	97	4.4	3782	3 US-09-861-289-4	Sequence 4, Appl1
36	97	4.4	3782	3 US-09-860-846-4	Sequence 4, Appl1
37	97	4.4	3782	3 US-09-988-384B-4	Sequence 4, Appl1
38	97	4.4	3782	3 US-09-836-821-4	Sequence 4, Appl1
39	97	4.4	3782	4 US-10-398-605-4	Sequence 4, Appl1
40	97	4.4	3782	4 US-10-271-889-47	Sequence 47, Appl
41	96.5	4.4	1595	4 US-10-425-115-238559	Sequence 238559,
42	96	4.4	421	4 US-10-092-900A-88	Sequence 88, Appl
43	96	4.4	442	4 US-10-425-115-223184	Sequence 223184,
44	96	4.4	711	4 US-10-425-114-46399	Sequence 46399, A
45	96	4.4	809	6 US-11-097-143-4965	Sequence 4965, Ap

ALIGNMENTS

RESULT 1									
US-10-617-038-29									
; Sequence 29, Application US/10617038									
; Publication No. US20040057963A1									
; GENERAL INFORMATION:									
; APPLICANT: Andersen, Peter									
; APPLICANT: Rosenkrands, Ida									
; APPLICANT: Strahm, Anette									
; TITLE OF INVENTION: Therapeutic TB Vaccine									
; FILE REFERENCE: SSI5AUSA									
; CURRENT APPLICATION NUMBER: US/10/617,038									
; CURRENT FILING DATE: 2003-07-11									
; PRIOR APPLICATION NUMBER: DK PA 2002 01098									
; PRIOR FILING DATE: 2002-07-13									
; PRIOR APPLICATION NUMBER: US 60/401,725									
; PRIOR FILING DATE: 2002-08-07									
; NUMBER OF SEQ ID NOS: 187									
; SOFTWARE: PatentIn version 3.2									
; SEQ ID NO 29									
; LENGTH: 413									
; TYPE: PRT									
; ORGANISM: Mycobacterium tuberculosis									
US-10-617-038-29									
Query Match 100.0%; Score 2195; DB 4; Length 413;									
Best Local Similarity 100.0%; Pred. No. 6.9e-207;									
Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MASSASDGTHERSAPRLSPVSGAMGPPMTGLVVAOSMRDYLGOQPKLIARPTAL	60						
DB	1	MASSASDGTHERSAPRLSPVSGAMGPPMTGLVVAOSMRDYLGOQPKLIARPTAL	60						
QY	61	AAQAARDEIVLGLKARBPVSNHRVFERISOEVAAGLEFYGNRMLEKSGFFAQPPLT	120						
DB	61	AAQAARDEIVLGLKARBPVSNHRVFERISOEVAAGLEFYGNRMLEKSGFFAQPPLT	120						
QY	121	EVAVRKVDORRRSPYRIFPDGFTPHPGPSQRLSTTANNREYALLRHEPPRWLYC	180						
DB	121	EVAVRKVDORRRSPYRIFPDGFTPHPGPSQRLSTTANNREYALLRHEPPRWLYC	180						
QY	181	VHGTMGRAPLDLAFAFMKLDHDELGLNIVMPVLPMHGRGGLPKGAFFPSEDVLDVH	240						
DB	181	VHGTMGRAPLDLAFAFMKLDHDELGLNIVMPVLPMHGRGGLPKGAFFPSEDVLDVH	240						
QY	241	GTAQAVMDIRRLSLWIRSOEBSLIGLNGLSIGGYIASLVASLEBGLACATIGVAVDLI	300						
DB	241	GTAQAVMDIRRLSLWIRSOEBSLIGLNGLSIGGYIASLVASLEBGLACATIGVAVDLI	300						
QY	301	ELAGHCGLRKQDPRRHVTAKAEPYGRMISPLSLTPLYPMGRFTYAGIADRLVHPRQV	360						
DB	301	ELAGHCGLRKQDPRRHVTAKAEPYGRMISPLSLTPLYPMGRFTYAGIADRLVHPRQV	360						

```

Db      301 ELLGRHCGLRHKDPRRHVTKMAEPIGSMISPLSLTFLVPMQGFIVAGIADRLVHPREGV 360
Oy      361 TRLMHWKGPETIVMTPGCHTGPFQSPRPVRFPQAALFEGGLDAPRTODRRA 413
       |||
Db      361 TRLMEHWKGPELIVWYPGGHTGPFQSPRPVRFPQAALFEGGLDAPRTODRRA 413

RESULT 2
US-10-425-115-286397
; Sequence 286397, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 286397
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_24291C.1.pep
US-10-425-115-286397

Query Match          6.7%, Score 146.5; DB 4; Length 366;
Best Local Similarity 24.6%; Pred.No.2.4e-05;
Matches 59; Conservative 31; Mismatches 81; Indels 69; Gaps 8;

Oy      208 NIWMPVL--PMHGPPQGQLPKGA-----VPGEEDVDVNGTAQAVMDIRRLSWIR 257
       || | | | : : : : : | : : : : | : : : : | : : : : | : : : :
Db      146 NIATMVLSPPYGGRRPMSQRGAKUCQVSDDLIGKATIDEA-----RSLLYWIQ 195
       || : : : : : | : : : : | : : : : | : : : : | : : : :

Oy      258 SQEEBSLILGNLSLGVIASLVASLE-----EGL----- 287
       :: : : : : : | : : : : | : : : : | : : : : | : : : :
Db      196 NBAGYGKKGICGLSNMGVHAAMVGSLHPTPVATLPFLAHSAVNPCEGYVKATAMDAL 255
       :: : : : : : | : : : : | : : : : | : : : : | : : : :

Oy      288 --ACAILGVPAVDLIELGRHCGLNKHKDPRRHVTVMKAEPISGLSIT----PLVPM 340
       - - : : | : : : : : | : : : : | : : : : | : : : : | : : : :
Db      256 RKDAVLTLDQDTLLAEADAQKSGITBOVR-----DLKSYLSLTDVTRFPVPKN 305
       - - : : | : : : : : | : : : : | : : : : | : : : : | : : : :

Oy      341 PGRFYGIADIIDLVAHQVTRLMEHWKPELIVWYPGGHTG--FFQSREPFRFQAALBO 398
       3 4 1 P G R F Y G I A D I L V A H Q V T R L M E H W K P E L I V W Y P G G H T G - - F F Q S R E P F R F Q A A L B O
Db      306 PQAVTFVGTADDGYIPHSVMELQKAMPSESVRWYTGGVSSFFLINDAFKAIYDALDR 365
       3 0 6 P Q A V T F V G T A D D G Y I P H S V M E L Q K A M P S E S V R W Y T G G V S S F F L I N D A F K A I Y D A L D R

RESULT 3
US-10-425-114-39123
; Sequence 39123, Application US/10425114
; Publication No. US2004003488BA1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 2003-04-28
; SEQ ID NO 39123
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
```

[illegible]

```
/ Publication No. US20040031072A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa Thomas J
/ APPLICANT: Kovacic David K
/ APPLICANT: Zhou Yihua
/ APPLICANT: Cao Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285664
/ SEQ ID NO 196342
/ LENGTH: 303
/ TYPE: PRT
/ ORGANISM: Glycine max
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(303)
/ OTHER INFORMATION: unsure at all Xaa locations
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_19323C.1.pap
US-10-424-599-196342

Query Match
Best Local Similarity 5.9%; Score 130; DB 4; Length 303;
Matches 66; Conservative 23; Mismatches 88; Indels 50; Gaps 12;

QY 203 DELGNTVVPVPMHGRPGQGLPKGAV-PPGEDVLDVHG-----TQAQVMDIRLLSMI 256
DB 88 EBOQLGALPVP--FYTARA PVLPGAXRCCVSDLT--LLGRPNLETPA-----RPLHW 137
QY 257 RSOEESLIGLNGSLSGYIASLVASLSEGLACALIGVPADLIBLGRHC----- 308
DB 138 CUDAGFSKMGGLGSLSGYHAAVMSLHP-----TVAVL-PLTSHSAVAVACESI 188
QY 309 LKH-----KDPRRHTVKA--EPIGMISPLSLT-----PLVPMGRFIVAGIADRL 353
DB 189 LKHGTAMEALRKDLAAQKVAAMTLEVERHRMNVLSLDTVRPPIPKNPVAVIFVATDDG 248
QY 354 VHPREQVRLMEHMKPELIVWYPGGHTGPF--QSRVRRFVQAALEQ 398
DB 249 YLPKHSVLELQKAMPGESEVRWYTGHSVFILHNDFFRAIKDGLDR 295

RESULT 6
US-11-097-143-27990
/ Sequence 27990, Application US/11097143
/ Publication No. US20050208558A1
/ GENERAL INFORMATION:
/ APPLICANT: Venter, J. Craig
/ APPLICANT: et al.
/ TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
/ TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
/ FILE REFERENCE: CL000728
/ CURRENT APPLICATION NUMBER: US/11/097,143
/ CURRENT FILING DATE: 2005-04-04
/ PRIOR APPLICATION NUMBER: 60/157,832
/ PRIOR FILING DATE: 1999-10-05
/ PRIOR APPLICATION NUMBER: 60/160,191
/ PRIOR FILING DATE: 1999-10-19
/ PRIOR APPLICATION NUMBER: 60/161,932
/ PRIOR FILING DATE: 1999-10-28
/ PRIOR APPLICATION NUMBER: 60/164,769
/ PRIOR FILING DATE: 1999-11-12
/ PRIOR APPLICATION NUMBER: 60/173,383
/ PRIOR FILING DATE: 1999-12-28
/ PRIOR APPLICATION NUMBER: 60/175,693
/ PRIOR FILING DATE: 2000-01-12
/ PRIOR APPLICATION NUMBER: 60/184,831
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: 60/191,637
```

```
/ PRIOR FILING DATE: 2000-03-23
/ NUMBER OF SEQ ID NOS: 43008
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 27990
/ LENGTH: 338
/ TYPE: PRT
/ ORGANISM: DROSOPHILA
US-11-097-143-27990

Query Match
Best Local Similarity 5.5%; Score 121; DB 6; Length 338;
Matches 68; Conservative 41; Mismatches 84; Indels 84; Gaps 16;

QY 165 YALLRHPEPR---PWLVCHGT--ENGRAPLDAVFRAMKLDLGLNTVMPVPMHG 218
DB 94 HAFVWTOPEERSKSPITLYFHNAGNMGH-----RMQNVWGIYHNLCHNVLMVEY----- 144
QY 219 PRGQGLPKGAVPPGEDVLDVHGTAQAVMDIRLLSMIRSGE--BSLIGLNGSLSGYI 276
DB 145 -RGYGLSTGV--PTGRGL-----VTDAARAALDYATRHDLDSQLIFGRSLGAV 192
QY 277 ASLVAS---LEBGLACALIGVPVADLIBLGRHCGLRHKDPRRHTVKAAPIGRMISPL- 332
DB 193 VVDVADTYTGQKLMCAIVENTFSSIPM-----AVELVHRAVKYIIPML 237
QY 333 -----SLTPLVPMGRFIV-AGIADRLVHPR-----EQYRLMEHMKPELIVWY 375
DB 238 FAKKYHSMKIKGKGVPLFTISGLADNLVPRPMRMALYTKGCSFIRLLE-----F 288
QY 376 PGGH-----TGFQSRPRRRFVQAALEQSGLDAP 405
DB 289 PGSHNDTWIVDGYQA--IGGFL-ABLQOQPLLKAP 322

RESULT 7
US-10-739-930-9602
/ Sequence 9602, Application US/10739930
/ Publication No. US20040216190A1
/ GENERAL INFORMATION:
/ APPLICANT: Kovacic, David K.
/ TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
/ TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
/ FILE REFERENCE: 38-21(53377)B
/ CURRENT APPLICATION NUMBER: US/10/739,930
/ CURRENT FILING DATE: 2003-12-18
/ NUMBER OF SEQ ID NOS: 11088
/ SEQ ID NO 9602
/ LENGTH: 322
/ TYPE: PRT
/ ORGANISM: Trifolium aestivum
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(322)
/ OTHER INFORMATION: unsure at all Xaa locations
/ FEATURE:
/ OTHER INFORMATION: Clone ID: TRIAB-23APR03-C123770_1.p
US-10-739-930-9602

Query Match
Best Local Similarity 5.4%; Score 118.5; DB 5; Length 322;
Matches 57; Conservative 25; Mismatches 76; Indels 67; Gaps 10;

QY 208 NIVMPVL--PMHGRPGQGLPKGA-----VPPGEDVLDVHGTAQAVMDIRLLSMIR 257
DB 130 NATVWLSFYGGARRPBMKGAKLQCVSDLLGKATIDDA-----RSLIYMDQ 179
QY 258 RSOEESLIGLNGSLSGYIASLVASLSEGLACALIGVP-----VADLIBLGRHC--- 307
DB 180 AEPGYKMGIGCLSGYHAAVMSL--LSTPIPTLFPLAHSVVPFCEGLYHATAM 236
QY 308 -GLRH-----KDPRRHTVKA--EPIGMISPLSLTLPVPMGRFIVAGIADRLVH 355
DB 237 BALREDAAALAKDATSLTEDAASGITIEOVXKQIT-----DGGYI 276
```



Qy 336 PEEQVTRLMEHWKPEIIVYPGHTGF--QSRPVRFLVQALLDQ 398  
Db 277 PHSYVLELOKAWPGSEFVRWVGSHVSSFLHNDSPFKALVDALDR 321

## RESULT 8

US-10-282-122A-62026  
; Sequence 62026, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Foreyeth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA, 034A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/491,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/506,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See file wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 62026  
; LENGTH: 574  
; TYPE: PRT  
; ORGANISM: Mycobacterium avium  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (199)..(199)  
; OTHER INFORMATION: X=any amino acid  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (310)..(310)  
; OTHER INFORMATION: X=any amino acid  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (349)..(349)  
; OTHER INFORMATION: X=any amino acid  
US-10-282-122A-62026

Query Match 5.1%; Score 112; DB 4; Length 574;  
Best Local Similarity 22.5%; Pred. No. 0.11;  
Matches 105; Conservative 46; Mismatches 148; Indels 168; Gaps 24;

Qy 23 SGAMGPFWHTGLY-----VAQSMRDYLGQPPDKLPAPRTALAAQAFRDEIVTLGLKA 76

Db 13 AGPLRPFADAGTPEADITHVAQRLTALTGEESDDRALA---VALLVRLRGSGVCVDLRA 69  
Qy 77 RRPVSNHRVFRISQEV-----AAG-----LEFYG-----NRRWL 106  
Db 70 -----VPAQVGAADLPMPAAGDWLAARASPLGPPPLRPFGLDLYFDRYWL 117  
Qy 107 EK-----PQGFPAQ-----PPLTEVAVARKVQDRRSFPIPFDSGFTPH 147  
Db 118 EEEQVCTDLALASAPAGVSESCYERLFPPEYEE-----ORAAARIIVSQALTVL 168  
Qy 148 GEPGQRLSTYANNREVALLL-----RHPEPRPMLVCVHTEMGA----- 189  
Db 169 GEPGTGK---TTVVARLLALVBOAERAGEPRPYIALAAPGKAARAEVAEIEHLD 225  
Qy 190 PLD---LAVFRAMKLDLGLNIVMPVLPMHGPRGQGLPKGAVPGEVDLDDVHTAQAV 246  
Db 226 PADRARLALGTLTHRLRIGRPDTSVAFKFN-RGNRLPHDI-----VVDETSNVSLTM 279  
Qy 247 WDRLRLSWIRSQEEESLIGLNGSLGYIASLVASLEEGLAAILGVPAVDLIELLGRH 306  
Db 280 --MARLEAVRPDTRLLVVG-----DDQLASVAGAX-----LADLYDGLAGR 321  
Qy 307 CGLRHKDRRTTVKMAE--IGRMISPLSLTPVPMGRFTYAGIADRLVHPRQVTRLM 364  
Db 322 AGVR-----VAALATPHRFGSALGALAA-----IRAADARVV----- 355  
Qy 365 EHWKPEIIVWYPGCHTGPFQS-RPVRRFVQA-----ALEQSGLLDA 404  
Db 366 -----ELLAAGGHIIEWVDSERPADRLREVLVSHALRLRSALILGA 396

## RESULT 9

US-10-437-963-183563  
; Sequence 183563, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 183563  
; LENGTH: 311  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: clone ID: PAT\_MRT4530\_80642C.1.pep  
US-10-437-963-183563

Query Match 5.1%; Score 111.5; DB 4; Length 311;  
Best Local Similarity 24.4%; Pred. No. 0.053;  
Matches 52; Conservative 26; Mismatches 68; Indels 67; Gaps 10;

Qy 208 NIYMPVL--PMHGRGQGLPKGA-----VFPGEVDLDDVHTAQAVMDIRLLSWIR 257  
Db 101 NIATWLESFYQGRRPMQHGSLQCVSDLLLGKATIDA-----RSLVYMQ 150  
Qy 258 SQEESLIGLNGSLGYIASLVASLEEGLAAILGVPAVDLIELLGRHGL----- 309  
Db 151 NEAGYKKIGICGLSMGVHAMVGSLLHP-----TPIATL-FLAPHSAAVVFCDGLY 201  
Qy 310 RH-----KDPKRNHTVVKMAEPIG-----RMISPLSLT-----PLVMP 341





CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO: 15631  
LENGTH: 742  
TYPE: PRT  
ORGANISM: Xanthomonas campestris  
US-10-369-493-15631

Query Match 4.8%; Score 104.5; DB 4; Length 742;  
Best Local Similarity 21.0%; Pred. No. 0.88;  
Matches 85; Conservative 50; Mismatches 161; Indels 109; Gaps 18;

QY 4 SASDGTHERSAPRLSPVLSGAMGPMHTGLVVAQSMRDVILGQPDKLPARPTIALAAQ 63  
DB 423 SASNGTHAAS-----FANNASVYVDTW-----SNTTTPQIB 454  
QY 64 AFRDEIVLLGLKARRPVSNNRVERISQEVAAGLEFYGNRRMLEKPSGFFA-----QPP 118  
DB 455 LFR-----ANGKTIATLLQNDVADPQHYPYAKYRQAPVBEFTLTAADGTP 501  
QY 119 LREAVAKYKDRRSFYRIFFDSGFTPHGEFGSQRLSYTANNREYALLRHPRPWL 178  
DB 502 L-----HYRLIKPAGFDP-----SKRY-----PVV 521  
QY 179 VCVHGTGEMGRAPLDLAVFRAMKLDHDELGINIVWPVLPMHGPRGQLP-KGAVPGEVDLD 237  
DB 522 VVYVGPPAAQTVDAMPGRGDALFDQYLAQGYVVFSLD--NRGTPRRGRAFG--ALY 576  
QY 238 DVHGTAAVMDIRRLSWIRSOE--EESLIGLNGSLGGYIA-SLVASLEBGLACAILGV 294  
DB 577 GKQGYIE-VDDQGVAVMLKTQKRWVDARIIGVGNNGSGYMTMLAKHSEAVACGVAGA 635  
QY 295 PVADLIELIGRRGGLRHKD-PRRHVTKMAP-IGRMISPLSTPLVPMGPRFTYAGIADR 352  
DB 636 PVTDW-GLYDTHYTERYMDLPAGNAGYREARIATHLDGLRAKL-----LIHGMADD 687  
QY 353 LV---HPRQVTRLMHEWKGKPEIWMYPGHGTGFFQSRPVRPFVQA 394  
DB 688 NVLFTSTALMSGLOQGTPELMTYPGAKHGSLGKTLALHRYXTA 732

Search completed: March 23, 2006, 06:28:19  
Job time : 88.8513 secs

This Page Blank (uspto)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 06:17:23 ; Search time 9.91096 Seconds  
(without alignments)  
1192.746 Million cell updates/sec

Title: US-10-617-038-29

Perfect score: 2195  
Sequence: 1 MASSASDCTHSRAFLSP.....AALEQSGILDAPRTGRDRA 413

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 169630 seqs, 28622889 residues

Total number of hits satisfying chosen parameters: 169630

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA\_New.\*  
1: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/1/pubppaa/PCT\_NEW\_PUB pep.\*  
5: /cgn2\_6/ptodata/1/pubppaa/US05\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB pep.\*  
7: /cgn2\_6/ptodata/1/pubppaa/US11\_NEW\_PUB pep.\*  
8: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	4.5	668	6 US-10-454-437-118	Sequence 118, Appl
2	96	4.4	2458	7 US-11-186-999-6	Sequence 6, Appl
3	95	4.3	300	7 US-11-179-977-4	Sequence 4, Appl
4	94.5	4.3	2455	7 US-11-186-999-4	Sequence 4, Appl
5	93	4.2	702	7 US-11-096-568A-14646	Sequence 14646, A
6	93	4.2	1096	6 US-10-995-561-710	Sequence 710, Appl
7	93	4.2	2458	7 US-11-186-999-13	Sequence 16978, A
8	92	4.2	402	7 US-11-096-568A-16978	Sequence 11, Appl
9	91	4.1	2458	7 US-11-186-999-11	Sequence 228, Appl
10	89.5	4.1	299	6 US-10-454-437-228	Sequence 9373, Ap
11	88	4.0	443	7 US-11-096-568A-9373	Sequence 9373, Ap
12	88	4.0	443	7 US-11-096-568A-9375	Sequence 1554, Ap
13	88	4.0	621	6 US-10-467-657-1554	Sequence 120, Appl
14	87.5	4.0	253	7 US-11-096-568A-33023	Sequence 33023, A
15	87.5	4.0	383	7 US-11-096-568A-12774	Sequence 12774, A
16	87.5	4.0	406	7 US-11-096-568A-18614	Sequence 18614, A
17	87.5	4.0	453	7 US-11-096-568A-18614	Sequence 18613, A
18	87.5	4.0	455	7 US-11-096-568A-18612	Sequence 18612, A
19	87.5	4.0	475	7 US-11-087-039-2233	Sequence 2233, Ap
20	86.5	3.9	563	6 US-10-454-437-120	Sequence 120, Appl
21	85.5	3.9	429	7 US-11-096-568A-33024	Sequence 33024, A
22	85	3.9	426	7 US-11-087-099-5180	Sequence 5180, Ap
23	85	3.9	2455	7 US-11-186-999-14	Sequence 14, Appl
24	85	3.9	2455	7 US-11-186-999-16	Sequence 16, Appl
25	84.5	3.8	383	7 US-11-096-568A-5747	Sequence 5747, Ap

26	84.5	3.8	2256	7 US-11-144-368-4	Sequence 4, Appl
27	84	3.8	298	7 US-11-179-977-6	Sequence 6, Appl
28	84	3.8	2348	6 US-10-450-224A-2	Sequence 2, Appl
29	83.5	3.8	264	7 US-11-096-568A-21268	Sequence 21268, A
30	83.5	3.8	297	7 US-11-096-568A-21267	Sequence 21267, A
31	83.5	3.8	674	7 US-11-167-048-1	Sequence 1, Appl
32	83	3.8	213	6 US-10-467-657-3394	Sequence 3394, Ap
33	82.5	3.8	366	6 US-10-329-258-27	Sequence 27, Appl
34	82.5	3.8	366	7 US-11-000-463-410	Sequence 410, App
35	82.5	3.8	366	7 US-11-000-463-882	Sequence 882, App
36	82.5	3.8	405	7 US-11-087-099-1566	Sequence 1566, Ap
37	82.5	3.8	407	7 US-11-087-099-5591	Sequence 5591, Ap
38	82.5	3.8	1057	6 US-10-330-773-253	Sequence 253, App
39	82	3.7	334	7 US-11-096-568A-10349	Sequence 10349, A
40	82	3.7	371	7 US-11-096-568A-10348	Sequence 10348, A
41	82	3.7	3073	7 US-11-143-980-50	Sequence 50, Appl
42	82	3.7	5712	7 US-11-143-980-47	Sequence 47, Appl
43	81.5	3.7	370	7 US-11-096-568A-28963	Sequence 28963, A
44	81.5	3.7	377	7 US-11-096-568A-28962	Sequence 28962, A
45	81.5	3.7	392	7 US-11-096-568A-29673	Sequence 29673, A

## ALIGNMENTS

RESULT 1  
US-10-454-437-118  
; Sequence 118, Application US/10454437  
; Publication No. US20050277115A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompeju, Markus  
; APPLICANT: Krogel, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zelder, Oskar  
; APPLICANT: Habberhauer, Gregor  
; TITLE OF INVENTION: CONYNEBACTERIUM GLUTAMINUM GENES ENCODING PROTEINS  
; FILE REFERENCE: BGI-128PCN  
; CURRENT APPLICATION NUMBER: US/10/454,437  
; PRIOR FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: US 60/141031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: DE 19931636.8  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19932125.6  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932126.4  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932127.2  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932128.0  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932129.9  
; PRIOR FILING DATE: 1999-07-19  
; PRIOR APPLICATION NUMBER: DE 19932226.0  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932920.6  
; PRIOR FILING DATE: 1999-07-14  
; PRIOR APPLICATION NUMBER: DE 19932922.2  
; PRIOR FILING DATE: 1999-07-14  
; Remaining Prior Application data removed - See file wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 442  
; SEQ ID NO 118  
; LENGTH: 668  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-10-454-437-118

Query Match 4.5%; Score 98; DB 6; Length 668;  
Best Local Similarity 20.8%; Pred. No. 0.59;  
Matches 91; Conservative 52; Mismatches 157; Indels 138; Gaps 25;  
QY 41 RDYIGQDPDKLPIARPTIALAAQAF-----RDEIVLVGLKRRPVSNHR- 84

```

Db      289 KEFLGCDPQVVFPPTESTSJQGLATTGNFLVLTILNNVSTEIVLPL--NDPTTEHH 346
Qy      85 --VFERSQEVAAGLEFYGNRRMLEKPSGFPAOPPELTV---AVRKYKDRRRSRFYRIE 138
Db      347 IDLPBHVAHVAVATSPLDGDEIWNQAS--FTEAPFLRAELPGALEAVKAPLQ----- 400
Qy      139 FDSGFTPHGEBSQSRWLSYTNANREYALLR--HPRRPMLVCHGTEBMGRAPLDLAV 195
Db      401 -----ENAGQETRHHWAT--SADGTXI PYFII GAFPEEPONTLVNHYG----- 441
Qy      196 FRWKHLDELGNITWPLVPMHGPR--RG--QGLPKGAVF-----PGEVDLVDVHGTA---- 243
Db      442 -----GFEVSLTPSHSPTRKIMLEKGYFVEANIRGGSEPGEBMSQATKIN 489
Qy      244 -QAVW-DIRRLTSW-I--RSQEEESLIGLNGLSIGYIAS-LVASTEEGLACAILGPVAD 298
Db      490 RMKWMEDHRAVLADLVERGYATPEQIATIGSGNGLITSGALTQYFEAGAAVVOYPLAD 549
Qy      299 LIELGRHCHLHKDPRRTTVK-----MAE-----PIGRMI-----SPLSLPLV--M 340
Db      550 ML-----RYHTWSAGASWMAEYGNPDPEBRAVIEQUSPVOAVGVGEKRTY 595
Qy      341 PGRPIYAGIADRLVHPREQVRLW-----EHMGKPELIWYPGCHTG-----F 382
Db      596 PPLATVTTSTRDVRVPAH--ARLFAQALIDAGQAVDVYENT-----GSHAGADNKKTA 648
Qy      383 FQSRVVRRFVQAALFQSG 400
Db      649 FVESLITWMEIKTLDOQG 666

```

```

RESULT 2
US-11-186-999-6
Sequence 6, Application US/11186999
Publication No. US20060019364A1
GENERAL INFORMATION:
APPLICANT: Bristol Myers Squibb Company
TITLE OF INVENTION: ACETYL COA CARBOXYLASE 2 SEQUENCES AND METHODS
FILE REFERENCE: BMS Dossier Number 10245 NP
CURRENT APPLICATION NUMBER: US/11/186,999
CURRENT FILING DATE: 2005-07-21
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn version 3.2
SEQ ID NO 6
LENGTH: 2458
TYPE: PRT
ORGANISM: Homo sapiens
US-11-186-999-6

Query Match          4.4% Score 96; DB 7; Length 2458;
Best Local Similarity 20.6%; Pred. No. 5,2; Indels 102; Gaps 23;
Matches 93; Conservative 46; Mismatches 131;

QY      37 AQSQRDYL--GQQPDKLPARPTLAALAAAFDEIVLLGL-----KRRPVSNRVER 88
Db      198 ALSLEAVLTGGEATRVVPTMRPSMS-----GLHLVKRGSHKKLDLRDP-- 242
        |||::|||::|
QY      89 ISQEVAAGLEF---YGNRRMLEK-----PSGFPAQPPPLTEVAVRKYKRSSFRIF-- 138
Db      243 ---TVASPAEFVTTRYGDRVIEKLIANGI-----AAVKCMRSIRRWAYEMFPNE 290
        |||::|||::|
QY      139 -----PDSGFTPHGPESQRMLSYTANNREYA-----TLT 169
Db      291 RAIFPVWWTEDLKANAETIKMADHYLVFGGP-----NNNYANVELIVDIAK 340
        |||::|||::|
QY      170 RHPEPRPW-----LYVCSTENGRAELDAVRAWTKHDELGINIV----- 210
Db      341 RIFRAVWAGMGHASENPKPELLCKNGVAFLGPSSE-----AMMALGDKIASVVAQTQL 396
        |||::|||::|
QY      211 MPVUPMHGP-----RGGGLPKG-AVFPGEDYLD-----DYH----- 240
Db      397 VFTLPWSGGSLTWEATEDDLQOGKRISVPBEDVDYDGCVKDVBEGLEAERIGFPLMIKAS 456
        |||::|||::|

```

[illegible]

```

RESULT 3
US-11-179-977-4
; Sequence 4, Application US/11179977
; Publication No. US20050245789A1
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Alpha/Beta Hydrolase-Fold Enzymes
; FILE REFERENCE: GC511-PCT
; CURRENT APPLICATION NUMBER: US/11/179,977
; CURRENT FILING DATE: 2005-07-12
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Bacillus
; PS-11-179-977-4

```

	Query Match	4.3%	Score 95;	DB 7;	Length 300;
	Best Local Similarity	23.1%	Pred. No. 0.38;		
	Matches	56;	Conservative 41;	Mismatches 109;	Indels 36; Gaps 11
Qy	171 HPEBRPMLVCHVGTENMGSRAPLDLAVFPAKMLHDEGLNIYVPLPMHGPRGQGLPKGAVF	230			
Db	78 HDPTNTIITIC HGVYTN---VLNSLKTMLFLDLGNVLLYDHRHRRGSG-GKTTSTYGF	131			
Qy	231 PGEDVLDVHGTAGTAAVMDIRRLLSWIRSQ-EESLSLIGLNGSLSGYIASLVAS--LEEGL	287			
Db	132 YEKD-----DLNKVSLLNKKTNRHGLIGHESMGAVTALTYAGAHCSQGA	178			
Qy	288 ACAILGVPAVDLIELGRHCGLRHKRPPRHITVKAERP-----GMISPLSLTPLV	339			
Db	179 DFYIADCPFCFDEQLAVRLAEAYRLPSMPLPLPIADFELKRGGRVAEVSPLAVIDKIE	238			
Qy	340 MPEGFYIAGLADRLVHREQYTRLMENMGKEIYWYR--GGHT-GFPSR-PVRFVQA	395			
Db	239 KPVLFHSHKDDYI--PVSSSTERLYEKKRGKPKALYIAENGHAMSYTNRHTYRTKTVQEF	296			
Qy	396 LE 397				
Db	297 LD 298				

```

RESULT 4
US-11-186-999-4
; Sequence 4, Application US/11186999
; Publication No. US20060019364A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: ACETYL COA CARBOXYLASE 2 SEQUENCES AND METHODS
; FILE REFERENCE: BMS Docket Number 10245 NP
; CURRENT APPLICATION NUMBER: US/11/186,999
; CURRENT FILING DATE: 2005-07-21
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 4
; LENGTH: 2455
; TYPE: PRT
; ORGANISM: Homo sapiens

```







```

Db      180 VASSSSSESTRKGRASLGLSLTLEAVLITLTTGEATRVPTMRSM--SGHLVLR-----GREH 233
QY      49 DKLPARP--TIALAAQAFRDEIVLGLKARRPVSNHRVFERI---SQEVAAGLEFYGNRR 104
Db      234 KLDLHRDPTVASPAF-----VTRXG-----GGRVIEKVLIAANGIAAVKCMSTIR 281
QY      105 WKEKSPGFAQPPPLTEVAVKVKDRRRSFYRIFFDSGFTPHGEGSQRMWLSYANNE 164
Db      282 MAYE---MFRNERAIFPVVWVTPEDLKANAEYIKMADHYVPVGGP-----NNNN 328
QY      165 YA-----LLLRHPRPRVLVCVHGTENGRAP-----LDLAVRAMKHLDEL 207
Db      329 YANVELIVDIARKIPYAXMKXXXXHASSEMPKLELLCKNGVAFGLPPESEAMALGDKIAS 388
QY      208 NIV-----MPVLPMPGP-----RGQGLPKG--AVFPGEDVLD-----DYH----- 240
Db      389 TVVAQVLQVPTLPMSSSGLTVENTEDDLOQKRISTPEVDYDGCYKVDDEGLAABERG 448
QY      241 -----GTAQAVMDIRLLSWIRSQEESLIGLNGLSGYIASLVASLE 284
Db      449 PKLMKASGGGKGIRKAKESADPPLPFQVQSEIRGSPILMKL-----AQHA 498
QY      285 EGLACILGVPAVDLIELGRHCGLRKDPRRRTVMAPRIGMISPLTLPVMPG-- 342
Db      499 RHLBVLQILDQYGNVAVSLFGRDCSIQ-----RRHQ-KIVEAPATYIPLAIFERMOCAR 553
QY      343 -----RFIYAGIADRLL-----VHPREQVTLMEH 366
Db      554 LAKTVGYSNKXVEYLYSQDGSFPHLEMLRLOY-----EH 589

RESULT 10
US-10-454-437-228
; Sequence 228, Application US/10454437
; Publication No. US20050277115A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroeger, Burkhard
; APPLICANT: Schroeder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Habernauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; FILE REFERENCE: BGI-128CPCN
; CURRENT APPLICATION NUMBER: US/10/454,437
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: DE 19931636.8
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932126.4
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932127.2
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932129.9
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: DE 19932226.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932920.6
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932922.2
; PRIOR FILING DATE: 1999-07-14
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 442
; SEQ ID NO 228
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum

```

```

US-10-454-437-228
Query Match      4.1%; Score 89.5; DB 6; Length 299;
Best Local Similarity 24.9%; Pred. No. 1.2;
Matches 63; Conservative 25; Mismatches 114; Indels 51; Gaps 10;

QY      176 PMLVCVHGTENGRAPDLAVFRAMKHLDELGINV-----MPVLPMPHGGGGLPKGAVF 230
Db      63 PCLVYFHGGGMSGGTLNMDATVHSL--VVGPIIATISVDYLAAPH----- 107
QY      231 PGSDVLDVHGTAAQVMDIRLLSWIRSQEESLIGLNGLSGYIASLVASLEEG----- 286
Db      108 PEPALDDAFVAVSAVLVDVSGLS-----IDSRVAIGDSAGSNIAAVTAQQLKRAVNG 162
QY      287 ---LACALIGVPVADLIE-----LGRHCGLRKDPRRRTVMAPRIGR---MISPL 332
Db      163 STPVLAAQVLIFPTVDVSTSTPSYLTFGKOCYCLTKDAMERYIEQYADGHDRTDPLSL 222
QY      333 SLTPVMPMGRTIYAGIADRLVHP-REOYTRLMHWKKEIYMYRGHTGFF-----Q 384
Db      223 LASDLSDEPPTIYVGECDVLAHEVRAYGOALLLEAGNSVTMTPEFKQIHAFINLGISSD 282
QY      385 SRPVRRFVQAAL 397
Db      283 ARPARRLRAELE 295

```

```

RESULT 11
US-11-096-568A-9373
; Sequence 9373, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 9373
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: misc. feature
; LOCATION: (1)-(443)
; OTHER INFORMATION: Ceres Seq. ID no. 14302452
US-11-096-568A-9373

```

```

Query Match      4.0%; Score 88; DB 7; Length 443;
Best Local Similarity 22.4%; Pred. No. 2.9;
Matches 83; Conservative 37; Mismatches 115; Indels 136; Gaps 21;

QY      44 LQQQDPKLPAPPTIALAAQAFRDEIVLGLK--ARRPVSNHRVFERISQEVAA----- 95
Db      11 VQGLGRILFGAS-----ARKAFHCIRLRGFPVQALAPVLENSSEPHLDPHANAAVFSLD 65
QY      96 --GLEFYGNRRWLEKPSGFFAQPPLTEVAV-----RKVKDRRSFYRIFFDSGFT 144
Db      66 EPGALALPTSRWFRTP-----PYLETAIVSVSRDGGCGHCGLRRLVGAVRVDVGE 119
QY      145 PHRPGSQRWLSYANNREYALLLRHPEPRPMLVCVHGTENGRAP-----LDLAVER 197
Db      120 WRDGRP-----VLRH-----SW-----TGIGRAELHVRVRVEADPRY 153
QY      198 AKMLHDELGLNIWMPVLPMPHPRGGGLPKGAVFPESVDLDVHGTAAQVMDIRLLSWIR 257
Db      154 VRFEDVVALN--PVVQLHGGASQ-----PIF-----SCKFIQVDR-----A 190
QY      258 SQEBSLIGLNGLSGYIASLVASLEEGIACALIGVPAVDLIELGRHCGLRKDPRRH 317
Db      191 SQPDH-----LGKTVSSSGSGEE-----KETENAG-----RRERKQW 224

```

[illegible]

```

RESULT 12
US-11-096-568A-9375
; Sequence 9375; Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-15928US2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 9375
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Trifolium aestivum
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(443)
; OTHER INFORMATION: Ceres Seq. ID no. 1661697
US-11-096-568A-9375

```

Query Match 4.0%; Score 88; DB 7; Length 443;  
 Best Local Similarity 22.4%; Pred. No. 2.9;  
 Matches 83; Conservative 37; Mismatches 115; Indels 136; Gaps 21.

```

Oy      44 GGGGDDXKPIAPFTIALAQAQRFDEVLGK---ARRPVSNRVERISQVAA----- 95
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      11 VGQLRLRLEPGAS-----ARKAPFCELRLEGFVQLAPVPLENSSEPHLDPHANAAPSLD 65
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      96 --GLEFYGNRMLEKPSGFPAQPPPLETAV-----RKQDRRSRFYRIFPDSGFT 144
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      66 EPGKALPSTRKFRPTE-----PYLELAIVYSRRDGGGHCYGLKRLVGAVRVNGVE 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      145 PHGPEGSGQRLWSTYANNEYALLNHPBPRLVYCVHGTENGRA-----LDLAIVR 197
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      120 WRDGRP-----VLRH-----SW-----TGIGRAELHVRVVEADPRY 153
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      198 AMKLHDELGLNIYMPVLPNHGPRGGCLPKGAVFPREGDVLDDVNGTQAQVMDIRLLSWIR 257
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      154 VRFDEYVALN---POVVOHLHGASQ-----PIF-----SCKFRIDVR-----A 190
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      258 SOEBSLGLNGLSLGGVYIASLVASLEGLACALIGVPAVDILELGRCHGJLHKQPRRH 317
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      191 SOPD-----LGKRYSSSGSSE-----KETEMG-----RRRKGM 224
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      318 TVKNAEPIGRMISPLS---TPLVPM-----PGRFYAGIADRLVHPRBOVT---RL 363
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      225 KVAIHDLSSGAVAAPFVPAVGSQDVARSNPQAMLI-----RADTSSSES 275
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      364 WEHMGKPEIIV 374
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      276 WQPMGRLE-AW 285
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 13  
US-10-467-657-1554  
Sequence 1554, Application US/10467657  
Publication No. US20050280581A1  
GENERAL INFORMATION:  
APPLICANT: CHIRON SpA  
APPLICANT: FONTANA Maria Rita  
APPLICANT: PIZZA Mariagrazia  
APPLICANT: MASIGNANI Vega

```

? APPLICANT: MONACI Elisabetha
? TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
? FILE REFERENCE:
? CURRENT APPLICATION NUMBER: US/10/467,657
? CURRENT FILING DATE: 2003-08-11
? PRIOR APPLICATION NUMBER: GB-0103424.8
? PRIOR FILING DATE: 2001-02-12
? NUMBER OF SEQ. ID NOS: 9218
? SOFTWARE: Seqwinn9, version 1.04
? SEQ ID NO 1554
? LENGTH: 621
? TYPE: PR1
? ORGANISM: Neisseria gonorrhoeae
? US-10-467-657-1554

```

Query Match	4.0%;	Score 88;	DB 6;	Length 621;
Best Local Similarity	25.2%;	Pred. No. 4.5;		
Matches	53;	Conservative	29;	Mismatches 74;
				Indels 54;
				Gaps 13;

```

Oy 211 MPTLPMH-GRQG-GLPKAVFPGSDYLDVNHGTA-----QAVMDLR-----LSW 255
Db 404 IPELPHYLSVGVKYLVEEGNAF-----VLNIRGGEGFGRMHQAAGISKKSVDLLAV 455
Oy 256 IRSQEEESL-----IGLNGLSGGYIASHVASLE-EGACALILGVPVADLIETL-----LGR 305
Db 460 VRDLSRGHSSPHHIGLGGSGNGSLITAAAPAREQSIGALVCEVPLTDMIRYPLSAGS 519
Oy 306 HCGLRHKDPRRHTVKMAEPIGMISPLS-----LPLVMPGRFTYAGIADRLVHP----- 356
Db 520 SMTDEYGNFQKY-----EACKRRRLGELSPPYHNLSGDIDYPALITTSLSDRVAPAHAK 574
Oy 357 -----REOYTRLMEHHGKPELWYVGGHTG 381
Db 575 FYAKLRETSPQSWLY--SPD-----GGGHTG 598

```

```

RESULT 14
US-11-096-568A-33023
; Sequence 33023, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Thebby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 33023
; LENGTH: 253
; TYPE: PR1
; ORGANISM: Arabidopsis thaliana
; FEATURES:
; NAME/KEY: misc.feature
; LOCATION: (1)..(253)
; OTHER INFORMATION: Ceres Seq. ID no. 13601781
US-11-096-568A-33023

```

Query Match	4.0%;	Score 87.5;	DB 7;	Length 253;
Best Local Similarity	21.8%;	Pred. No. 1.5;		
Matches 43;	Conservative 28;	Mismatches 73;	Indels 53;	Gaps 8;

[illegible]

```

QY      285 EGLCAILGVPVADLIE 301
          | | : : | :
Db      144 VDAVVGFGYGTSPSELAD 160

```

```

RESULT 15
US-11-096-568A-12774
: Sequence 12774, Application US/11096568A
: Publication No. US20060048240A1
GENERAL INFORMATION:
APPLICANT: Alexandrov, Nikolai et al.
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
TITLE OF INVENTION: Theory
FILE REFERENCE: 2750-1592PUS2
CURRENT APPLICATION NUMBER: US/11/096,568A
CURRENT FILING DATE: 2005-04-01
NUMBER OF SEQ ID NOS: 34471
SEQ ID NO 12774
LENGTH: 383
TYPE: PRT
ORGANISM: Triticum aestivum
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(383)
OTHER INFORMATION: Ceres Seq. ID no. 14303558
US-11-096-568A-12774

```

Query Match	4.0%;	Score 87.5;	DB 7;	Length 383;
Best Local Similarity	23.2%;	Pred. No. 2.6;		
Matches	78;	Conservative	46;	Mismatches 137;
			Indels	75;
			Gaps	19;

```
Qy      89 ISEVVAAGLEFYENRMRLBKPSGGFGFQOPPLTEVVAVRKVKRRRSFYIFEDSGFTBPg 148
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      4 VASTVAARFAFP-----PSPSYGVPERPPSAAAA-----DGAIVELSG 44

Qy      149 EPG----SQRMSTYANRREYALLRHPRRWLYCVHG--TEMGRAPLDAVRAMKH 202
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      45 VPREGVEARRLETTKRGSEVMAMYAKOPEARLTLLYSHGNAADJGOM-YELFV---ELS 99

Qy      203 DEGLNLIWMPVLPMHGPBGQLPKGAFVEGEDVDLDVGTAQAAMDRIRLSMWSIOEE 262
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      100 SHLNATLM-----GYDSYGQGSSGKRSE---QNTVADIEMAY--RCIETRYGASEE 146

Qy      263 SLGLNGLSLGCTIAILVASLESBGLACALISPVADVILLGRCHGLRHDPRRHTYYMA 322
       ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      147 NII-LTGQSVSGSGETILDASLRPHLRPAVALVHSPISS-----GLKWTVPVGHYTWF- 195

Qy      323 EPIGRMISPSLTFLVPMGFRTIYGIADRLV---HPREQVTRLME-HMGKPBIWTYPG 378
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      196 ----DIYNKDINKVALKCFFVLVIH-GTDADDVDCSHGR-----ALMELSKVCYEPLAWXKG 246

Qy      379 HTGFPOSRP-----VARPYQALLSOGSL--DAPT 407
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      247 NHCNLELVPFIYIHMLKKFI-TAIEKSPPKLKSDSPS 281
```

Search completed: March 23, 2006, 06:29:40  
Job time : 10.911 secs

**This Page Blank (usp)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 04:56:29 ; Search time 31.904 Seconds  
(without alignments)  
1652.628 Million cell updates/sec

Title: US-10-617-038-30

Perfect score: 637  
Sequence: 1 MSTRPRRSHGIRAVGPYAWA.....DMPAYVIGEHLSVETAVAV 120

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_21.\*  
1: geneseqp1980a:\*  
2: geneseqp1990a:\*  
3: geneseqp2000a:\*  
4: geneseqp2001a:\*  
5: geneseqp2002a:\*  
6: geneseqp2003a:\*  
7: geneseqp2003b:\*  
8: geneseqp2004a:\*  
9: geneseqp2005a:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	637	100.0	120	8	AD137309
2	80	12.6	551	8	ADK68544
3	79.5	12.5	468	2	AAW71633
4	78	12.2	310	7	ABO79202
5	77.5	12.2	456	6	ABU40175
6	75	11.8	510	7	ABO76512
7	75	11.8	594	8	ADG22732
8	75	11.8	8805	4	ABBE7112
9	74.5	11.7	223	8	AD065060
10	74	11.6	330	7	ABO77614
11	74	11.6	535	6	ABU34179
12	73	11.5	69	4	ABB39158
13	73	11.5	69	4	AAW32650
14	73	11.5	69	4	AAW72398
15	73	11.5	69	4	AAW59811
16	73	11.5	69	4	ABG54088
17	73	11.5	69	5	ABG42216
18	73	11.5	258	4	AAW75597
19	73	11.5	361	8	ADJ67604
20	73	11.5	449	7	ADK64595
21	73	11.5	715	4	AAU34857
22	73	11.5	715	6	ABU47919
23	73	11.5	715	6	ABU15299
24	73	11.5	715	8	ADG45230

25	73	11.5	734	4	AA95634	AA95634	Human pro
26	73	11.5	766	7	ADB7917	ADB7917	Human put
27	73	11.5	766	8	ABW81904	ABW81904	Tumour-as
28	73	11.5	766	8	ADS88275	ADS88275	Human pro
29	73	11.5	1277	4	ABB11784	ABB11784	Human sec
30	73	11.5	1277	8	ADS12057	ADS12057	Human the
31	73	11.5	1277	9	AEA20840	AEA20840	Novel hum
32	73	11.5	1277	9	AEA20840	AEA20840	Novel hum
33	72.5	11.4	221	6	AAU51687	AAU51687	Human ULP
34	72.5	11.4	221	6	ABM48206	ABM48206	Propionib
35	72.5	11.4	492	4	ABG14735	ABG14735	Novel hum
36	72.5	11.3	195	9	AEA79725	AEA79725	S. griseu
37	72	11.3	195	9	AEA48218	AEA48218	Streptoc
38	72	11.3	255	9	AEA80310	AEA80310	S. griseu
39	72	11.3	255	9	AEA48814	AEA48814	Mature st
40	72	11.3	457	9	AEA80324	AEA80324	Streptomy
41	72	11.3	457	9	AEA48827	AEA48827	Streptoc
42	72	11.3	510	2	AAW49835	AAW49835	Thermus a
43	71.5	11.2	759	6	ABP78096	ABP78096	N. gonorr
44	71	11.1	292	2	AAW88406	AAW88406	Trichoder
45	70.5	11.1	268	8	ADG73564	ADG73564	Glycopept

#### ALIGNMENTS

RESULT 1  
ID AD137309 standard; protein; 120 AA.  
AC AD137309;  
DT 22-APR-2004 (first entry)  
DE M. tuberculosis low oxygen induced antigen Rv2628 SEQ ID NO:30.  
KW mycobacterial infection; vaccine; tuberculosis;  
KW Mycobacterium tuberculosis; immunisation; antibacterial; gene therapy;  
KW low oxygen induced antigen.  
OS Mycobacterium tuberculosis.  
PN WO2004006952-A2.  
PD 22-JAN-2004.  
PF 08-UTL-2003; 2003WO-DK000477.  
PR 13-UTL-2002; 2002DK-00001098.  
PI (STAT-) STATENS SERUM INST.  
PI Andersen P, Rosenkrands I, Stryhn A;  
DR WPI; 2004-122778/12.  
DR N-PSDB; AD137354.  
PT Use of one or more polypeptides or their fragments, which are expressed during the latent stage of the mycobacterial infection, and/or nucleic acid encoding the polypeptides, for a therapeutic vaccine against tuberculosis.  
PS Claim 3; SEQ ID NO 30; 76pp; English.  
XX The present invention describes polypeptides or their fragments, which are expressed during the latent stage of a mycobacterial infection, and/or nucleic acids encoding the polypeptides, which are useful for creating a therapeutic vaccine against tuberculosis. Also described: (1) a therapeutic vaccine against tuberculosis comprising one or more polypeptides; (2) a method for treating an animal, including a human being, with tuberculosis caused by virulent mycobacteria, e.g. by Mycobacterium tuberculosis, M. africanum or M. bovis; (3) a method for immunising an animal, including a human being, against tuberculosis

CC caused by virulent mycobacteria; (4) a method of diagnosing tuberculosis  
CC caused by virulent mycobacteria in an animal, including a human being;  
CC (5) a method for diagnosing previous or ongoing infection with a virulent  
CC mycobacterium; and (6) a method of diagnosing mycobacterium tuberculosis  
CC infection in a subject. The polypeptides have antibacterial activities,  
CC and can be used in vaccines and in gene therapy. The polypeptides are  
CC useful for the manufacture of a therapeutic vaccine for treating an  
CC individual who is infected by a virulent mycobacterium, e.g. M.  
CC tuberculosis, and who is not vaccinated with BCG against tuberculosis.  
CC The present sequence represents a low oxygen induced antigen, which is  
CC used in the exemplification of the present invention.

SO Sequence 120 AA;

Query Match 100.0%; Score 637; DB 8; Length 120;  
Best Local Similarity 100.0%; Pred. No. 5, 1e-69;  
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MGTORRHSGIRAVGYAMAGRCGRIGRGVHQEAMNNLAIMPRVQSAITYQVDRSH 60  
DB 1 MGTORRHSGIRAVGYAMAGRCGRIGRGVHQEAMNNLAIMPRVQSAITYQVDRSH 60

OY 61 DGRITAVPGDEITSTVSGMLSELGTQSPADELARAVRIGDMPAAVAIGEHLSVEIAVAV 120  
DB 61 DGRITAVPGDEITSTVSGMLSELGTQSPADELARAVRIGDMPAAVAIGEHLSVEIAVAV 120

RESULT 2  
ADK68544  
ID ADK68544 standard; protein; 551 AA.

AC ADK68544;  
DT 21-APR-2005 (first entry)

DE Plant full length insert polypeptide seqid 39387.

KM plant protectant; plant growth regulant; gene therapy; plant;  
KM recombinant DNA construct; physical array; plant breeding marker;  
KM cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KM extreme osmotic condition; pathogen tolerance; pest tolerance;  
KM growth rate; cell cycle pathway; disease resistance;  
KM galactomannan production; lignin production; plant growth regulator;  
KM yield; plant growth; plant development; seed oil; protein yield;  
KM protein content.

XX Unidentified.

OS US2004034888-A1.

PN 19-FEB-2004.

PD 28-APR-2003; 2003US-0042514.

PF 06-MAY-1999; 99US-00304517.

PR 05-NOV-2001; 2001US-00985678.

XX (LIUJ/) LIU J.

PA (ZHOU/) ZHOU Y.

PA (KOVA/) KOVALIC D K.

PA (SCRE/) SCREEN S E.

PA (TABAS/) TABASKA J E.

PA (CAOY/) CAO Y.

XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;

DR WPI; 2004-180133/17.

XX New recombinant DNA construct, useful for improving plant tolerance to

PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or

PT pests, for conferring increased resistance to plant disease, or for

PS Claim 1; SEQ ID NO 39387; 15pp; English.

XX The invention describes a recombinant DNA construct comprising a  
XX polynucleotide consisting of a sequence encoding an amino acid sequence  
XX available in electronic form from the US patent office at  
XX ftp://seqdata.uspto.gov/sequence.html?docID:2004034888. The polynucleotide  
XX of the invention are also useful in physical arrays of molecules and as  
XX plant breeding markers. The recombinant DNA construct is useful for  
XX improving plant tolerance to cold, heat, drought, herbicides, extreme  
XX osmotic conditions, pathogens or pests, for manipulating growth rate in  
XX plant cells by modification of the cell cycle pathway, for conferring  
XX increased resistance to plant disease, for producing galactomannan,  
XX lignin or plant growth regulators, for increasing the rate of homologous  
XX recombination in plants, for improving yield by modification of  
XX photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
XX or by providing improved plant growth and development under at least one  
XX stress condition or for modifying seed oil or protein yield and/or  
XX content. This is the amino acid sequence of a plant full length insert  
XX polypeptide that can be used in the recombinant DNA construct of the  
XX invention.

SO Sequence 551 AA;

Query Match 12.6%; Score 80; DB 8; Length 551;  
Best Local Similarity 27.0%; Pred. No. 1.9; Indels 64; Gaps 10;  
Matches 43; Conservative 15; Mismatches 37;

OY 6 PRH-----SGIRAVGYAMAGRCG-----RIGRWGVHQEAMNNLAIMPRKY----- 47  
DB 71 PRHLPHGPRRRRRYRABQPHL-HGRGGLREVVRAGEPQVNH--LHL-----PRVPLPR 123

OY 48 -----QSATTY-----QVDRSHDGRITAVPGDEITSTVSGMLSELGTQSP 88  
DB 124 PLGQHQLGRRRAGLYGMPRGHNLHQGTEREHVGSRRGLPRPR-----QLGSGVP 173

OY 89 LADELARAVRIGDMPAAVAIGE-----HLVSEIAVA 119  
DB 174 QGANHARRAVG--PALVELGETEVAERPRRVGVEHVA 210

RESULT 3  
AAW71633  
ID AAW71633 standard; protein; 468 AA.

AC AAW71633;

DT 26-NOV-1998 (first entry)

DE Omega-cyclohexane fatty acid biosynthesis enzyme #1 ORF1.

XX Omega-cyclohexane fatty acid; biosynthesis; enzyme; detection;

XX Allcyclobacillus acidocaldarius ATCC 27009; identification; microbe.

OS Allcyclobacillus acidocaldarius.

PN JP10234376-A.

PD 08-SEP-1998.

PF 28-FEB-1997; 97JP-00046570.

PR 28-FEB-1997; 97JP-00046570.

PA (KIRI ) KIRIN BEVERAGE KK.

DR WPI; 1998-535030/46.

XX N-PSDB; AAV58229.

XX New nucleic acid - useful for detection and identification of genus

PT Allcyclobacillus microorganism(s).

PS Claim 5; Page 23-25; 37pp; Japanese.

CC The present sequence represents a protein from ORF1 of a new nucleic acid  
CC which encodes enzymes which participate in the biosynthesis of Omega-  
CC cyclohexane fatty acid. The nucleic acid is isolated from  
CC Alicyclobacillus acidocaldarius. The present invention also describes  
CC primers and probes containing all or part of the nucleic acid from  
CC Alicyclobacillus acidocaldarius. The primers and probes may be used for  
CC detection and/or identification of a microorganism of genus  
CC Alicyclobacillus. The method can detect and identify Alicyclobacillus  
CC genus rapidly and easily  
XX

SO Sequence 468 AA;

Query Match 12.5%; Score 79.5; DB 2; Length 468;  
Best Local Similarity 28.3%; Pred. No. 1.8;  
Matches 30; Conservative 11; Mismatches 40; Indels 25; Gaps 6;

QY 24 GRIGRGVGHQAMNNLAIMHPRKVSATIVQVTDNR-----SHDRTARVPDELTSTV--- 76  
DB 368 GSTRVAVRQRE-KORLGV--NVPVNAIMETIQDGVRIERDRBELVFPADQVLCCTGOL 423  
77 ---SGWLSEI-----GTQSPLADELARAVRIGDWPAAVAIG 109  
DB 424 PGRTATETKLPPIHVRVHIVGAGDSRDINARARREA-WMAAYDIG 468

RESULT 4  
ABO79202  
ID ABO79202 standard; protein; 310 AA.

AC ABO79202;

DT 29-JUL-2004 (first entry)

DE Pseudomonas aeruginosa polypeptide #11377.

KM Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.

OS Pseudomonas aeruginosa.

PN US651795-B1.

PD 22-APR-2003.

PF 18-FEB-1999; 99US-00252991.

PR 18-FEB-1998; 98US-0074788P.

PR 27-JUL-1998; 98US-0094190P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;

DR WPI; 2003-615309/58.

DR N-PSDB; ABD12773.

PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.

XX disclosure; SEQ ID NO 27948; 455bp; English.

PS The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polymers encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biochip technology. Sequences ABO67826-

CC ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html)  
XX

SO Sequence 310 AA;

Query Match 12.2%; Score 78; DB 7; Length 310;  
Best Local Similarity 26.0%; Pred. No. 1.6;  
Matches 27; Conservative 4; Mismatches 33; Indels 40; Gaps 3;

QY 9 SGIRAVGPYAMAGRCGRIGRMVGHQAMNNLAIMHPRKVSATIVQVTDNRSHDGRPARVP 68  
DB 237 AGRRCAAGACAGACAGRRG-----DGGTRKTA 263  
QY 69 GDEITSTVSGWLSELTQSPPLADELARAVRIGDWPAAVAIGHL 112  
DB 264 GERQAPPACTWVRQRLG---QCRQETARRRLRRQWPA---GGHL 300

RESULT 5  
ABU40175  
ID ABU40175 standard; protein; 456 AA.

AC ABU40175;

DT 19-JUN-2003 (first entry)

DE Protein encoded by prokaryotic essential gene #25702.

KM Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Pseudomonas putida.

PN MO200277183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

PA (ELIT-) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW,  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

DR WPI; 2003-029926/02.

DR N-PSDB; ACA44045.

PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 68099; 1766bp; English.

PS The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)





CC for preparing enzymes for synthesis of D-bases (D = 2,6-diaminopurine),  
 CC particularly D, dUMP and dTMP, or polynucleotides containing these bases,  
 CC polymerases involved in metabolism of D-bases and deoxynucleotide  
 CC analogs, for chemotherapy of AIDS. The genes, and encoded polypeptides,  
 CC can be used for detection and/or identification of S-2L, and for  
 CC identifying agents that modulate synthesis of D-bases or polynucleotides  
 CC containing them, and fusions of S-2L polypeptides with an antigen can be  
 CC used to raise specific antibodies, useful for detecting S-2L. This  
 CC sequence corresponds to one of the proteins encoded by the cyanophage S-  
 CC 2L genome.

XX  
 SQ Sequence 594 AA;

Query Match 11.8%; Score 75; DB 8; Length 594;  
 Best Local Similarity 31.2%; Pred. No. 8.7;  
 Matches 25; Conservative 9; Mismatches 28; Indels 18; Gaps 4;

QY 58 RSHDGR-ARVPGDEITVSGWLSL-----GTGSPFLADELARAV-----RIGDM 102  
 DB 19 RHHGGRHCGDGGEPGRSIHEWAGSIALGHGTQAPIDSELPRIICPPADRLERIGG 78

QY 103 PAAVAI---GEHLSVEIAVA 119  
 DB 79 PEAADAGRPEGHVLTVOVS 98

RESULT 8  
 ABB67112  
 ID ABB67112 standard; protein; 8805 AA.

XX ABB67112;  
 AC ABB67112;  
 DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 28128.

XX Drosophila; developmental biology; cell signalling; insecticide;  
 XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL11215.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.

XX Disclosure; SEQ ID NO 28128; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX  
 SQ Sequence 8805 AA;

Query Match 11.8%; Score 75; DB 4; Length 8805;  
 Best Local Similarity 29.9%; Pred. No. 2.9e+02;  
 Matches 29; Conservative 15; Mismatches 41; Indels 12; Gaps 4;

QY 28 RWGHOEAMMNLAIWHPKQVATITVYQVTDG8-HDGRFARV-----PDEITVSGWL 80  
 DB 2326 RGLVDEQLIKIV--DPTTKQPISTVEAIARSIDYDPKTIIDSEGPVDLITATKGL 2383

QY 81 SELGTQSPFLADELARAVRIGDMPAVAIGEHLSEIA 117  
 DB 2384 SVVGAFLVAABGALRTVRFVTDPT--GEQIPVEVA 2417

RESULT 9  
 ADO65060  
 ID ADO65060 standard; protein; 223 AA.

XX ADO65060;  
 AC ADO65060;

XX 07-OCT-2004 (first entry)

XX Novel human protein sequence #33.

XX osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;  
 XX gene therapy; diagnostic marker; morbid state; osteoporosis;  
 XX neurological disease; Alzheimer's disease; Parkinson's disease; dementia;  
 XX cancer.

XX Homo sapiens.

XX EP1440981-A2.

XX 28-JUL-2004.

XX 21-JAN-2004; 2004EP-00001196.

XX 21-JAN-2003; 2003JP-00102206.

XX 09-MAY-2003; 2003JP-00113392.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 PI Yamamoto J, Isono Y, Nagai K, Irie R;

XX WPI; 2004-535376/52.

XX N-PSDB; ADO62872.

XX Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,  
 PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.

XX Claim 1; SEQ ID NO 2221; 2449pp; English.

XX The invention relates to 2495 novel polynucleotides (I) and their encoded  
 CC polypeptides, sequences hybridizing to these nucleotides, sequences  
 CC encoding partial polypeptides and sequences having 70% or 90% identity to  
 CC the nucleotide and protein sequences. The nucleotides and polypeptides  
 CC are useful as diagnostic markers or therapeutic target for the diseases  
 CC or morbid states. They are also useful for treating osteoporosis,  
 CC neurological diseases, Alzheimer's diseases, Parkinson's diseases,  
 CC dementia and various cancers. This sequence corresponds to a protein  
 CC sequence of the invention.

XX Sequence 223 AA;

Query Match 11.7%; Score 74.5; DB 8; Length 223;  
 Best Local Similarity 27.3%; Pred. No. 2.8;  
 Matches 35; Conservative 13; Mismatches 29; Indels 51; Gaps 9;

QY 15 GPYANAGRCGRIGWGHQEAAMNLAIW-----HPRKVOGA-----TTIVQVT 56  
 ||:| || ||||| ||:|

Db 60 GEMALGR-RRIGR-----LALMAAPRRSGPRTSEVGSRRHGMFWRSEOS 108  
QY 57 DRSHDR-TARVGDDEITSTVSG--W--LSELTGTSPLA-----DELAR----- 95  
Db 109 PPARSGRGTVQVPGAIVSGVPCPTRSAVPGOGERPFLARGRTAEGSGEGERGVVVG 168  
QY 96 -AVRIGDW 102  
Db 169 AALRVGTW 176

RESULT 10  
AB077614  
ID AB077614 standard; protein; 330 AA.  
XX  
AC AB077614;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Pseudomonas aeruginosa polypeptide #9789.  
XX  
KM Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN US651795-B1.  
XX  
PD 22-APR-2003.  
XX  
PF 18-FEB-1999; 99US-00252991.  
XX  
PR 18-FEB-1998; 98US-0074788P.  
XX  
PR 27-JUL-1998; 98US-0094190P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
DR WPI; 2003-615309/58.  
XX  
DR N-PSDB; ABD11185.  
XX  
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
XX  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX  
PT pathological conditions resulting from bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 26360; 455pp; English.  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnosis and  
XX therapy of pathological conditions, as molecular targets for diagnostics,  
XX prophylaxis and treatment of pathological conditions resulting from a  
XX bacterial infection, for evaluating a compound, such as a polypeptide,  
XX for the ability to bind a P. aeruginosa nucleic acid, as components of  
XX effective antibacterial targets, as targets for antibacterial drugs,  
XX including anti-P. aeruginosa drugs, as templates for recombinant  
XX production of P. aeruginosa-derived peptides or polypeptides, as target  
XX components for diagnosis and/or treatment of P. aeruginosa-caused  
XX infection, and in detection of P. aeruginosa sequences or other sequences  
XX of Pseudomonas species using biochip technology. Sequences AB067826-  
XX AB084396 represent P. aeruginosa polypeptides of the invention. Note: The  
XX sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format from USPTO at  
XX seqdata.uspto.gov/sequence.html  
SQ Sequence 330 AA;  
Query Match 11.6%; Score 74; DB 7; Length 330;  
Best Local Similarity 25.2%; Pred. No. 5.3;  
Matches 31; Conservative 13; Mismatches 59; Indels 20; Gaps 4;  
QY 2 STGRPHSGIRAVGPAAW-----GRCGRIGW-----GHOEAMMLAIW 43  
Db 26 TSRRRPGPLPMAAPVMAARRLGRSGRSRSGRWPSRVCTTSMPTRAOSSSAMIATGA 85

QY 44 PRKVSATITYQVTDSS-HDGRARVPGDEITSTVSGWLSELTGTSPLADELARAVRIGDW 102  
Db 86 RRLRTISPAWSAYPPSAQSRCSITSSAVCSRSSGCVRLTAQPPCTPR-RRCRLRRRG 144  
QY 103 PAA 105  
Db 145 PAA 147

RESULT 11  
ABU34179  
ID ABU34179 standard; protein; 535 AA.  
XX  
AC ABU34179;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #19706.  
XX  
KM Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Mycobacterium avium.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002MO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
XX  
PR 06-SEP-2001; 2001US-00948993.  
XX  
PR 25-OCT-2001; 2001US-0342923P.  
XX  
PR 08-FEB-2002; 2002US-00072851.  
XX  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselebeck R, Ohlsen XL, Zyskind JW;  
XX  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
XX  
DR N-PSDB; ACA38049.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
XX  
PT for homologous nucleic acids required for cellular proliferation to  
XX  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 62103; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 623 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the  
XX proliferation of an organism. The antisense nucleic acids are useful for  
XX identifying proteins or screening for homologous nucleic acids required



```
Db          2 ELSDVSGGSKATP-----ASTANSDVATIPDTPLKENEGRVKTADPAKSEISKIIE 57
QY          114 VEIA 117
           |::|
Db          58 VQVA 61

RESULT 14
AAM72398
ID AAM72398 standard; protein; 69 AA.
AC AAM72398;
XX
XX
DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 32704.
XX
XX DE Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; Leukemia; Lymphoma; myeloma.
XX
XX OS Homo sapiens.
XX
XX PN WO200157276-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000668.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-488900/53.
XX
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human bone marrow.
XX
XX PS Example 4; SEQ ID NO 32704; 658bp + Sequence Listing; English.
XX
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukemia and myeloma. The present sequence is a
XX CC protein encoded by one of the probes of the invention
XX
XX SQ Sequence 69 AA;

Query Match          11.5%; Score 73; DB 4; Length 69;
Best Local Similarity 28.1%; Pred. No. 0.92;
Matches 18; Conservative 15; Mismatches 27; Indels 4; Gaps 1;

QY 54 QVTRSHDGRTRAVPGDEITVSGWLSGLTQSPDLADELARAARVIGDMPAAVAIGEHLS 113
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 2 ELSDVSGGSKATP-----ASTANSDVATIPDTPLKENEGRVKTADPAKSEISKIIE 57

QY 114 VEIA 117
   |::|
Db 58 VQVA 61

RESULT 15
AAM59811
ID AAM59811 standard; protein; 69 AA.
XX
```

```
AC AAM59811;
XX
XX DT 05-NOV-2001 (first entry)
XX
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 31916.
XX
XX DE Human; brain expressed exon; gene expression analysis; probe; microarray;
XX KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX
XX OS Homo sapiens.
XX
XX PN WO200157275-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000667.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-483446/52.
XX
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains.
XX
XX PS Example 4; SEQ ID NO 31916; 650bp + Sequence Listing; English.
XX
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC brain. They can be used to measure gene expression in brain cell samples,
XX CC which may enable the diagnosis and improved treatment of nervous system
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX CC epilepsy and cancer. The present sequence is a protein encoded by one of
XX CC the probes of the invention
XX
XX SQ Sequence 69 AA;

Query Match          11.5%; Score 73; DB 4; Length 69;
Best Local Similarity 28.1%; Pred. No. 0.92;
Matches 18; Conservative 15; Mismatches 27; Indels 4; Gaps 1;

QY 54 QVTRSHDGRTRAVPGDEITVSGWLSGLTQSPDLADELARAARVIGDMPAAVAIGEHLS 113
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 2 ELSDVSGGSKATP-----ASTANSDVATIPDTPLKENEGRVKTADPAKSEISKIIE 57

QY 114 VEIA 117
   |::|
Db 58 VQVA 61

Search completed: March 23, 2006, 05:10:58
Job time : 33.904 secs
```



A:Residues: 1-691 <TET>  
A:Cross-references: UNIPROT:Q9K0B2, UNI-PARC:UPI00000C4543, GB:AE002425, GB:AE002098, NIT  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NM80702  
C:Superfamily: competence protein ComEC

Query Match 13.4%; Score 85.5; DB 2; Length 691;  
Best Local Similarity 27.1%; Pred. No. 1.1;  
Matches 29; Conservative 14; Mismatches 51; Indels 13; Gaps 3;

Oy PRHSGRVAVGPYAMGRCRIGRMGVCHOEAMNNLAIWPRKY-----OSATYQ 54  
Db PQRSSVLMLAAFAFMWRGRSLSAMTWQALAAVILFPPLAVLGVTGLSGVAALIA 300  
+ : : ||| | : : : : : : : : : : : : : : : : : :  
55 VTRSDHG-RTARVPGEDEITSVS-GWLSELTGSPLADELARAVRI 99  
+ : : ||| | : : : : : : : : : : : : : : : : : :  
301 CSGRHLEGKRQTALRGWAASVLSLVLLGYLPASFLISPLVNAAI 347

RESULT 3  
B81937  
competence protein MMA0906 [imported] - Neisseria meningitidis (strain Z2491 serogroup A  
C:Species: Neisseria meningitidis  
C>Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 09-Jul-2004  
C:Accession: B81937  
R.Packhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
J.; Holtroyd, S.; Jagella, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A>Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: A81775, PMID:2022256; PMID:10761919  
A:Accession: B81937  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-691 <PAR>  
A:Cross-references: UNIPROT:Q9UYB8, UNI-PARC:UPI00000C4942, GB:AL162754, GB:AL157959; NIT  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: comA, MMA0906  
C:Superfamily: competence protein ComEC

Query Match 13.1%; Score 83.5; DB 2; Length 691;  
Best Local Similarity 26.2%; Pred. No. 1.8;  
Matches 28; Conservative 15; Mismatches 51; Indels 13; Gaps 3;

Oy PRHSGRVAVGPYAMGRCRIGRMGVCHOEAMNNLAIWPRKY-----OSATYQ 54  
Db PQRSSVLMLAAFAFMWRGRSLSAMTWQALAAVILFPPLAVLGVTGLSGVAALIA 300  
+ : : ||| | : : : : : : : : : : : : : : : : : :  
241 PTQSSVLMLAAFAFMWRGRSLSAMTWQALAAVILFPPLAVLGVTGLSGVAALIA 300  
+ : : ||| | : : : : : : : : : : : : : : : : : :  
Oy 55 VTRSDHG-RTARVPGEDEITSVS-GWLSELTGSPLADELARAVRI 99  
+ : : ||| | : : : : : : : : : : : : : : : : : :  
Db 301 CSGRHLEGKRQTALRGWAASVLSLVLLGYLPASFLISPLVNAAI 347

RESULT 4  
G98145  
probable aminotransferase y4ub [imported] - Agrobacterium tumefaciens (strain C58, Cerec  
C:Species: Agrobacterium tumefaciens  
C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 31-Dec-2004  
C:Accession: G98145  
R.Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, W.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A87359; PMID:21608551; PMID:11743194  
A:Accession: G98145  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-520 <KUB>  
A:Cross-references: UNIPROT:Q8U6P6, UNI-PARC:UPI00000D26BF, GB:AE007870, PIDN:AAK86689.1  
C:Genetics:  
A:Gene: AGR\_L\_242  
A:Map position: linear chromosome

```

Query Match          12.6%; Score 80.5; DB 2; Length 520;
Best Local Similarity 28.0%; Pred. No.2.6;
Matches      28; Conservative    9; Mismatches   56; Indels    7; Gaps     2;

QY      2 STORPRHSGIRAVPYVMAGRCGRIGRWGCHQEMMNLAIWHPRKVQSATIIYQVTRSDH 61
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      8 SPEEPPFSA--AGNNQGLAARARVMASMNISNPMSASALWLPBRISPMTIIDIKNAEMD 65
       : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      62 GRIVRVPEDEITSTVSGMLSELGTQSPLADELARAVRGD 101
       :: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      66 RNAAVHFPTQLKDFPASKLGE-----FTIVETGKGIRICD 100
       : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
S39867
competence protein comA - Neisseria gonorrhoeae
C:Species: Neisseria gonorrhoeae
C:date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C:Accession: S39867; S71019
R:Factus, D.; Meyer, T.F.
Mol. Microbiol. 10, 699-712, 1993
A>Title: A novel determinant (comA) essential for natural transformation competence in N
A:Reference number: S39867; MUID:95020543; PMID:7934834
A:Accession: S39867
A:Molecule type: DNA
A:Residues: 1-691 <FAC>
A:Cross-references: UNIPROT:P51973; UNIPARC:UP10000127EB83; GB:S75480; NID:g913747; PIND:
R:Fusenberger, M.; Factus, D.; Meiler, J.; Meyer, T.F.
Mol. Microbiol. 19, 1095-1105, 1996
A>Title: A novel peptidoglycan-linked lipoprotein (ComL) that functions in natural trans
A:Reference number: S71019; MUID:96249702; PMID:8830266
A:Accession: S71019
A:Molecule type: DNA
A:Residues: 644-691 <FUS>
A:Cross-references: UNIPARC:UP100001796F5; EMBL:Z49895; NID:g1107832
A:Experimental source: strain MS11
C:Genetics:
A:Gene: comA
C:Superfamily: competence protein ComEC
C:Keywords: transmembrane protein; transport protein
F:183-203/Domain: transmembrane #status predicted <TM1>
F:220-240/Domain: transmembrane #status predicted <TM2>
F:280-300/Domain: transmembrane #status predicted <TM3>
F:323-342/Domain: transmembrane #status predicted <TM4>
F:347-367/Domain: transmembrane #status predicted <TM5>
F:396-416/Domain: transmembrane #status predicted <TM6>

Query Match          12.5%; Score 79.5; DB 2; Length 691;
Best Local Similarity 26.2%; Pred. No.4.5;
Matches      28; Conservative   14; Mismatches   52; Indels   13; Gaps     3;

QY      6 PRHSGITAVCPYVMAGRCGRIGRWGCHQEMMNLAIWHPKV'-----QSATTIYO 54
       : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      241 PTQRSVMLLAFAVAAMRRGRLSMATWMQALAVALFLFDPLAVGVGTWISFGVLALIVA 300
       : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      55 VTDRSHDG-RTARVPGEITSTVS-GWLSELTGQSPLADELARAVRI 99
       :: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      301 CAGELTEGKQTAVRGQAASVLSVLILGYLFASLPVLSPLVANVISI 347
       : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
G87496
peptidase, M23/M37 family [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C:Accession: G87496
R:Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.B.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Dukkin, A.S.; Gattm, M.L.; Haft, D.H.; Kolon
N. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A>Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647

```







Db 686 EPIADORAALQYVIDEY 702

RESULT 9

JC5958

sperm surface protein - human  
C:Species: Homo sapiens (man)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004

C:Accession: JC5958  
R:Shankar, S.; Mohapatra, B.; Suri, A.  
Biochem. Biophys. Res. Commun. 243, 561-565, 1998

A:Title: Cloning of a novel human testis mRNA specifically expressed in testicular haplo

A:Reference number: JC5958; MUID:98153804; PMID:9480848

A:Accession: JC5958

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-766 <SHA>

A:Cross-references: UNIPROT:O60905; UNIPARC:UPI000017C354; EMBL:X91879

Query Match 11.5%; Score 73; DB 2; Length 766;

Best Local Similarity 28.1%; Pred. No. 23;

Matches 18; Conservative 15; Mismatches 27; Indels 4; Gaps 1;

QY 54 QYTDPSHQRTRAVPDEITSTVSGWLSLGTQSPDLADLARAARIGDWPAAVAIGEHLS 113

Db 84 ELSDVSGGSKATTP---ASTANSQVATIPDTPLKENEGFKVTDAPNKSEISKHLE 139

QY 114 VEIA 117

Db 140 VOVA 143

RESULT 10

S57953

C4BP protein alpha chain precursor - rat

C:Species: Rattus norvegicus (Norway rat)  
C:Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 09-Jul-2004

C:Accession: S57953

R:Hillarp, A.; Thern, A.; Dahlback, B.  
submitted to the EMBL Data Library, July 1995

A:Title: Molecular cloning of rat C4b-binding protein alpha- and beta-chains: stru

A:Reference number: S57953

A:Accession: S57953

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-558 <HL>

A:Cross-references: UNIPROT:O63514; UNIPARC:UPI0000126C29; EMBL:Z50051; NID:G899379; PTD

C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology

F:15-72/Domain: complement factor H repeat homology <FH1>

F:77-134/Domain: complement factor H repeat homology <FH2>

F:139-199/Domain: complement factor H repeat homology <FH3>

F:204-258/Domain: complement factor H repeat homology <FH4>

F:263-324/Domain: complement factor H repeat homology <FH5>

F:328-386/Domain: complement factor H repeat homology <FH6>

F:390-443/Domain: complement factor H repeat homology <FH7>

F:447-501/Domain: complement factor H repeat homology <FH8>

Query Match 11.4%; Score 72.5; DB 2; Length 558;

Best Local Similarity 29.9%; Pred. No. 18;

Matches 26; Conservative 11; Mismatches 37; Indels 13; Gaps 4;

QY 23 CGRIGRW-----GVHQEAMNLAIMPRKYGASATY---QYTDPSHQRTRAVPDEITS 74

Db 373 CKSDGTHMRTSPCHQSCFPALAHGRYTKSSVYRVQVTECEBG--YLVGE---A 427

QY 75 TVSGWLSLGTQSPDLADLARAARIGD 101

Db 428 TISCWYSQWTPAAPQCKALCRKPEIGN 454

RESULT 11

A53669 streptogrisin C (EC 3.4.21.-) precursor - Streptomyces griseus

C:Species: Streptomyces griseus

C:Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 05-Oct-2004

C:Accession: A53669

R:Sidhu, S.S.; Kaimar, G.B.; Willis, L.G.; Borgford, T.J.

J. Biol. Chem. 269, 20167-20171, 1994

A:Title: Streptomyces griseus protease C, A novel enzyme of the chymotrypsin superfamily

A:Reference number: A53669; MUID:94327573; PMID:8051104

A:Accession: A53669

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-457 <SID>

A:Cross-references: UNIPROT:P52320; UNIPARC:UPI00000139F; GB:L29018; NID:G532654; PIDN...

Query Match 11.3%; Score 72; DB 2; Length 457;

Best Local Similarity 28.0%; Pred. No. 16;

Matches 26; Conservative 14; Mismatches 27; Indels 26; Gaps 5;

QY 17 YAMAGRCRIGRW--GVHQEA-----MMNLAIMPRKYGASATTYQVDR 58

Db 233 FATAGHGRVGTGTTNGVQAQGTGSGTFPGRIAWATNMTPRRLVNG--YGRBDV 290

QY 59 SHDGRTRAVPDEI--TSTVSGW---LSEIGT 85

Db 291 TVAGSTASVAGSVCRSSTTGWMCRTIQOLNT 323

RESULT 12

JC2294

carboxypeptidase (EC 3.4.-.-) Tag - Thermus aquaticus (strain YT-1)

C:Species: Thermus aquaticus

C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 05-Oct-2004

C:Accession: JC2294; PC2202

R:Lee, S.H.; Taguchi, H.; Yoshimura, E.; Mingawa, E.; Kamnogawa, S.; Ohta, T.; Matsumi

Biosci. Biotechnol. Biochem. 58, 1490-1495, 1994

A:Title: Carboxypeptidase Tag, a thermostable zinc enzyme, from Thermus aquaticus YT-1: '

A:Reference number: JC2294; MUID:94369103; PMID:7765282

A:Accession: JC2294

A:Molecule type: DNA

A:Residues: 1-511 <LEB1>

A:Cross-references: UNIPROT:P42663; UNIPARC:UPI00001285FD; DDBJ:D17669; NID:9497631; PID:

A:Accession: PC2202

A:Molecule type: protein

A:Residues: 1-511 <LEB2>

A:Cross-references: UNIPARC:UPI00001285FD

C:Comment: This enzyme hydrolyzes the peptide bond at the C-terminus of peptides and pro

C:Comment: This enzyme is a metallo-carboxypeptidase with thermostability and a broad su

A:Gene: Tag

C:Superfamily: Zn-dependent carboxypeptidase, Tag type

Query Match 11.3%; Score 72; DB 2; Length 511;

Best Local Similarity 27.2%; Pred. No. 18;

Matches 31; Conservative 14; Mismatches 37; Indels 32; Gaps 6;

QY 18 AMAGRCRIGRWGVHQEAMNLAIMPRKYGASATTYQT-----DRSHGRTARVPGD 70

Db 18 AYLSGLGLAW--DQRTM-----PRKGHNRARQVALARLLHERATPR----- 62

QY 71 EITSTVSGWLSLGTQSPDLADLARAARIGDWPAAV---AIGEHLSVEIATA 119

Db 63 -----IGWLEKVEGSLVEDPLSDAIVNVAAMRAVARRAATPERLAVELAAQ 111

RESULT 13

G70932

probable monooxygenase cytochrome P450 Rv0568 - Mycobacterium tuberculosis (strain H37Rv)

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Mycobacterium tuberculosis

CjDate: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004

CjAccession: G70932

RiCole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G. A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome A:Reference number: A70500; MUID:9829587; PMID:9634230

A:Accession: G70932

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-472 <COL>

A:Cross-references: UNIPROT:O53765; UNIPARC:UPI0000126BB9; GB:AL021942; GB:AL123456; NIT

A:Experimental source: strain H37Rv

CjGenetics:

A:Gene: RV0568

CjSuperfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase

P:258-410/Domain: cytochrome P450 homology <P45>

F:388/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 11.1%; Score 71; DB 2; Length 472;  
Best Local Similarity 21.0%; Pred. NO. 21;  
Matches 37; Conservative 23; Mismatches 46; Indels 70; Gaps 9;

OY 7 RHSGIRAVGYAMAGRCGRIGRW---GVHQEAMNLAIW--PRKVQ--SATY-QVTR 58  
Db 165 RLAAIRKVP-----RLNIVGPWATLALNPSLNNRLMSLRRIREADALLVAEIDR 219  
OY 59 -----SHDGTAR-----VPGEITSTVSGWSELQ 86  
Db 220 RADPDLAARTDTLAMLVRABEDGRTMERELDQLITLLVAGHDTATVATGLSMALERTLR 279  
OY 87 SPLADELAAVRIGD-----WPAAYAIGEH,SYEIAVA 119  
Db 280 HPVT--LAKAVQADADSAAGDPAGDEYLDANAKETLRPFVYVDGRVLTAEVYA 333

RESULT 14

AF0666

probable exported protein STY1442 (imported) - Salmonella enterica subsp. enterica serov

C:Species: Salmonella enterica subsp. enterica serovar Typh

A:Note: this species has also been called Salmonella typh

CjDate: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

CjAccession: AF0666

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, S.; Moutle, S.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, S. Nature 413, 848-852, 2001

A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AF0666

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-541 <PAR>

A:Cross-references: UNIPARC:UPI0000059EB8; GB:AL513382; PIDN:CAD01703.1; PID:G16502553;

CjGenetics:

A:Gene: STY1442

CjSuperfamily: periplasmic glucans biosynthesis protein mdg

Query Match 11.1%; Score 71; DB 2; Length 541;  
Best Local Similarity 22.9%; Pred. NO. 24;  
Matches 22; Conservative 11; Mismatches 33; Indels 30; Gaps 3;

OY 28 RWGVRQEAAMNL-----AIWPRKVASATYQVTRSHDGTARVPGEITSTV 76  
Db 359 KWGKAVASLMEPTTGETLDMNVVCFWQPEKAIA-----GDILAFNY 400  
OY 77 S-GWSELQTSPLADELAAVRIGDWPAAVAIGH 111  
Db 401 RLWMSAQPPVQSPILRAVMAITTKMGFPEDGAPGHH 436

[illegible]

Search completed: March 23, 2006, 05:26:37  
Job time : 5.77423 secs

**THIS PAGE BLANK (see 10)**

**This page Blank (usps 10)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 04:56:43 ; Search time 30.6915 Seconds  
(without alignments)  
2758.529 Million cell updates/sec

Title: US-10-617-038-30

Perfect score: 637  
Sequence: 1 MSTOPRHSRGIRAVGPYMA.....DMPAAVIGELHSLVEIAVAV 120

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Uniprot\_05.80:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	637	100.0	120	006184_MYCTU	006184 mycobacteri
2	622	97.6	120	077Y70_MYCBO	077Y70 mycobacteri
3	260	40.8	87	073MD2_MYCBO	073MD2 mycobacteri
4	87.5	13.7	447	0511J1_SILPO	0511J1 silicibacte
5	85.5	13.4	691	09KOB2_NEIMB	09KOB2 neisseria m
6	83.5	13.1	691	09JVB8_NEIMA	09JVB8 neisseria m
7	82.5	13.0	331	05YS68_NOCFA	05YS68 nocardia fa
8	82	12.9	281	07VVV9_BORBP	07VVV9 bordetella
9	81	12.7	277	07WKL0_BORBR	07WKL0 bordetella
10	80.5	12.6	263	08ZTE5_PYPAB	08ZTE5 pyrobaculum
11	80.5	12.6	520	07CCT0_AGRIS	07CCT0 agrobacteri
12	79.5	12.5	691	1 COMA_NEIGO	P51973 neisseria g
13	79	12.4	270	05KXV7_GEOXA	05KXV7 geobacillus
14	78	12.2	163	093MF2_ARYMD	093MF2 amycolactops
15	77.5	12.2	456	088E24_PSRBP	088E24 pseudomonas
16	77	12.1	277	07N269_CHRYO	07N269 chromobace
17	77	12.1	316	07W2W1_BORPA	07W2W1 bordetella
18	76.5	12.0	283	04KE05_PSEBS	04KE05 pseudomonas
19	76	11.9	455	04K5Q9_PSEBS	04K5Q9 pseudomonas
20	76	11.9	757	05GZM7_XANOR	05GZM7 xanthomonas
21	75.5	11.9	316	07WDM1_BORBR	07WDM1 bordetella
22	75.5	11.9	510	05SLM3_THET8	05SLM3 thermus the
23	75.5	11.9	510	07ZG13_THET8	07ZG13 thermus the
24	75	11.8	337	04RMT2_TETNG	04RMT2 tetradion n
25	75	11.8	8805	07KR73_DROME	07KR73 drosophila
26	74.5	11.7	223	06ZMG9_HUMAN	06ZMG9 homo sapien
27	74.5	11.7	252	05Z1Y0_NOCFA	05Z1Y0 nocardia fa
28	74.5	11.7	527	08BSN9_PSEBS	08BSN9 pseudomonas
29	74.5	11.7	731	08FMS1_CORER	08FMS1 corynebacte
30	74	11.6	284	082OD7_STRAW	082OD7 streptomyce
31	74	11.6	609	09A6T7_CAUCR	09A6T7 caulobacter

32	73.5	11.5	214	2	05U2U1_RAT	05U2U1 rattus norv
33	73.5	11.5	223	2	067U16_ORISA	067U16 oryza sativ
34	73.5	11.5	316	2	054MF6_DICDI	054MF6 dictyosteli
35	73.5	11.5	407	2	09KYV1_STRCO	09KYV1 streptomyce
36	73	11.5	173	2	0889T1_PSEBS	0889T1 pseudomonas
37	73	11.5	204	2	066WC7_HUMAN	066WC7 homo sapien
38	73	11.5	365	2	08RH77_ECOLA	08RH77 escherichia
39	73	11.5	712	2	070J33_MACFA	070J33 macaca fasc
40	73	11.5	715	1	PDHP_ECOLI	P07658 escherichia
41	73	11.5	715	2	05PJ14_SALPA	05PJ14 salmonella
42	73	11.5	715	2	08PAX4_ECOL6	08PAX4 escherichia
43	73	11.5	715	2	08ZKE7_SALTY	08ZKE7 salmonella
44	73	11.5	715	2	082100_SALTI	082100 salmonella
45	73	11.5	734	2	09H811_HUMAN	09H811 homo sapien

#### ALIGNMENTS

RESULT 1  
ID 006184 MYCTU PRELIMINARY; PRT; 120 AA.  
AC 006184\_07D6V2;  
DT 01-JUL-1997 (TRMBLrel. 04, Created)  
DT 01-JUL-1997 (TRMBLrel. 04, Last sequence update)  
DT 13-SEP-2005 (TRMBLrel. 31, Last annotation update)  
DE Hypochemical protein.  
GN OrderedAccessionNames=MT2703, RV2628;  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinomycetales;  
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium;  
OX NCBI\_TaxID=1773;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230; DOI=10.1038/31159;  
RA Cole S.T., Brosch R., Parkhill J., Garfield T., Churcher C.M.,  
RA Harris D.E., Gordon S.V., Eigmler K., Gas S., Barry C.E. III,  
RA Tekala F., Badcock K., Basham D., Brown D., Chillingworth T.,  
RA Connor R., Davies R.M., Devlin K., Felwell T., Gentles S., Hamlin N.,  
RA Holtroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S.,  
RA Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream W.A.,  
RA Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulten J.E., Taylor K., Whitehead S., Barrett B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence";  
RL Nature 393:537-544 (1998).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=CDC 1551 / Oshkosh;  
RX MEDLINE=22206494; PubMed=12218036;  
RA DOI=10.1126/DB.184.19.5479-5490.2002;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J.D., DeBoy R.T., Dodson R.J., Gwinn M.L., Haft D.H.,  
RA Hickey E.K., Kolony J.F., Nelson W.C., Umayam L.A., Brimble M.D.,  
RA Salzberg S.L., Delcher A., Utterback T.R., Weidman J.F., Khouri H.M.,  
RA Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,  
RA Frazer C.M.;  
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains";  
RL J. Bacteriol. 184:5479-5490 (2002).  
DR EMBL: BX842580; CAB0635.1; -: Genomic DNA.  
DR EMBL: AE000516; AAK47019.1; -: Genomic DNA.  
DR PIR: C70573; C70573.  
DR TIGR: MT2703; -;  
DR TIGR: MT2703; -;  
KW Complete proteome; Hypochemical protein.  
SQ SEQUENCE 120 AA; 13129 MM; CAAAGC42A61B42F0 CRC64;  
Query Match 100.0%; Score 637; DB 2; Length 120;  
Best Local Similarity 100.0%; Pred. No. 2,le-55;  
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 MSTORPRHSGIRAVGPYANAGRCGRIGRWGVHGOEAMNMLAIWHPKRVGSATTIYQVTDNRSH 60
Db 1 MSTORPRHSGIRAVGPYANAGRCGRIGRWGVHGOEAMNMLAIWHPKRVGSATTIYQVTDNRSH 60
Qy 61 DORTARVPDEITSTVSGWLSLGTQSPPLADELARAVRIGDMPAAVAIGEHLVSEIYAV 120
Db 61 DORTARVPDEITSTVSGWLSLGTQSPPLADELARAVRIGDMPAAVAIGEHLVSEIYAV 120

RESULT 2
Q7TY70_MYCBO PRELIMINARY; PRT; 120 AA.
ID Q7TY70_MYCBO PRELIMINARY; PRT; 120 AA.
AC Q7TY70;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein Mb2661.
GN OrderedLOCusNames=MB2661;
OC Mycobacterium bovis.
OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium tuberculosis complex.
OC NCBI_TaxId=1765;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;
RA Garrier T., Eiglmietter K., Camus J.-C., Medina N., Mancoor H.,
RA Pryor M., Dutfoy S., Grondin S., Lacroix C., Monsenpe C., Simon S.,
RA Harth B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewison R.G.,
RT "The complete genome sequence of Mycobacterium bovis."
RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
DR EMBL: BX248343; CAD94846.1; -; Genomic_DNA.
KM Complete proteome.
SQ SEQUENCE 120 AA; 13214 MW; 54287642A589EB28 CRC64;

Query Match 97.6%; Score 622; DB 2; Length 120;
Best Local Similarity 98.3%; Pred. No. 6.6e-54;
Matches 118; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSTORPRHSGIRAVGPYANAGRCGRIGRWGVHGOEAMNMLAIWHPKRVGSATTIYQVTDNRSH 60
Db 1 MSTORPRHSGIRAVGPYANAGRCGRIGRWGVHGOEAMNMLAIWHPKRVGSATTIYQVTDNRSH 60
Qy 61 DORTARVPDEITSTVSGWLSLGTQSPPLADELARAVRIGDMPAAVAIGEHLVSEIYAV 120
Db 61 DORTARVPDEITSTVSGWLSLGTQSPPLADELARAVRIGDMPAAVAIGEHLVSEIYAV 120

RESULT 3
Q73WD2_MYCPA PRELIMINARY; PRT; 87 AA.
ID Q73WD2_MYCPA PRELIMINARY; PRT; 87 AA.
AC Q73WD2;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLOCusNames=MAP2728;
OC Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium avium complex (MAC).
OC NCBI_TaxId=1770;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=k10;
RA Li L., Bannanting J., Zhang Q., Amonsai A., Alt D., Kapur V.;
RA Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE017237; AAS05045.1; -; Genomic_DNA.
KM Complete proteome; Hypothetical protein.
SQ SEQUENCE 87 AA; 9576 MW; 684CD5A0BD21034A CRC64;
```

```
Query Match 40.8%; Score 260; DB 2; Length 87;
Best Local Similarity 57.6%; Pred. No. 4.7e-18;
Matches 53; Conservative 12; Mismatches 17; Indels 10; Gaps 2;

Qy 28 RWGVHGOEAMNMLAIWHPKRVGSATTIYQVTDNRSHOGRTRARVPDEITSTVSGWLSLGTQSP 87
Db 5 KMHVH-----W-PGRPRATTIYQVTDNRSHOGRTRARVPDEITSTVSGWLSLGTQSP 54

Qy 88 PLADELARAVRIGDMPAAVAIGEHLVSEIYAV 119
Db 55 PLVDELERAVRIGDMPAAVAIGEHLVSEIYAV 86

RESULT 4
Q5ILJ1_SILPO PRELIMINARY; PRT; 447 AA.
ID Q5ILJ1_SILPO PRELIMINARY; PRT; 447 AA.
AC Q5ILJ1;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE NADH:ubiquinone oxidoreductase, Na(+)-translocating, A subunit
DE (EC 1.6.5.-).
GN Name=nqra; OrderedLOCusNames=SP0A0028;
OC Silicibacter pomeroyi.
OC Bacteria; Proteobacteriia; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Silicibacter.
OC NCBI_TaxId=89184;
RN [1]
RP NUCLEOTIDE SEQUENCE (LARGE SCALE GENOMIC DNA).
RC STRAIN=DS-3 / ATCC 700808 / DSM 15171;
RX PubMed=15602564; DOI=10.1038/nature03170;
RA Moran M.A., Buchan A., Gonzalez J.M., Heidelberg J.F., Whitman W.B.,
RA Kliene R.P., Henriksen J.R., King G.M., Belas R., Fugua C.,
RA Brinkac L.M., Lewis M., Jost S., Weaver B., Pai G., Eisen J.A.,
RA Rane E., Shelton W.M., Ye W., Miller T.R., Carlton J., Rasko D.A.,
RA Paulsen I.T., Ren Q., Daugherty S.C., DeBoy R.T., Dodson R.J.,
RA Haft D.H., Selengut J., Ward N.;
RT "Genome sequence of Silicibacter pomeroyi reveals adaptations to the
RT marine environment".
RL Nature 432:910-913 (2004).
DR EMBL: CP000032; AAV97168.1; -; Genomic_DNA.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0016655; F:oxidoreductase activity, acting on NADH or . . .; IEA.
DR GO: GO:0006814; P:sodium ion transport; IEA.
DR InterPro: IPR008703; NQRA.
DR Pfam: PF05896; NQRA; 1.
DR TIGFams: TIGR01936; nqra; 1.
KM Complete proteome; Oxidoreductase; Plasmid; Ubiquinone.
SQ SEQUENCE 447 AA; 47132 MW; BD7AD132AAB25EF6 CRC64;

Query Match 13.7%; Score 87.5; DB 2; Length 447;
Best Local Similarity 30.1%; Pred. No. 4;
Matches 34; Conservative 15; Mismatches 43; Indels 21; Gaps 5;

Qy 9 SGIRAVGPYANAGRCGRIGRWGVHGOEAMNMLAIWHPKRVGSATTIYQVTDNRSH 56
Db 307 SGRQAGGFALVGRARQVTL-IEEDRKQHTLPLPISFVAQPVLSFSLVLT 365

Qy 57 DRSHDRTARVPDEITSTVSGWLSLGTQSPPLADELARAVRIGDMPAAVAIGEHLVSEIYAV 109
Db 366 SNLNGRRAMP-----TGTFEELMPQVYLPOLLRLALVMDTDPAQLG 410

RESULT 5
Q9KOB2_NEIMB PRELIMINARY; PRT; 691 AA.
ID Q9KOB2_NEIMB PRELIMINARY; PRT; 691 AA.
AC Q9KOB2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
```

```

DE Competence protein ComA.
GN Name=comA; OrderedLocustNames=NM00702;
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxId=491;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MC58 / Serogroup B;
MEDLINE=20175755; PubMed=10710307; DOI=10.1126/science.287.5459.1809;
RA Tetteijn H., Saunders N.J., Heidelberg J.P., Jeffries A.C.,
RA Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F.,
RA Dodson R.J., Nelson W.C., Gwinn M.L., Deboy R.T., Peterson J.D.,
RA Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D.,
RA Doucherty B.A., Mason T.M., Ciecko A., Parksey D.S., Blair E.,
RA Citron H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H.M.,
RA Qin H., Vamathevan J.J., Gill J., Scarlato V., Maignani V., Pizzo M.,
RA Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R.,
RA Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
DR EMBL; AE002098; AAF4119.1; -; Genomic_DNA.
DR PIR; G81167; G81167.
DR TIGR; NM00702; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0030420; P:establishment of competence for transformation; IEA.
DR InterPro; IPR001279; Blactmase-like.
DR InterPro; IPR004477; ComEC_N-term.
DR InterPro; IPR004797; ComEC_Rec2.
DR Pfam; PF03772; Competence; 1.
DR Pfam; PF00753; Lactamase_B; 1.
DR TIGRFAMs; TIGR00360; ComEC_N-term; 1.
DR TIGRFAMs; TIGR00361; ComEC_Rec2; 1.
KW Complete proteome.
SQ SEQUENCE 691 AA; 74408 MW; 5EC00CE0B166C6FE CRC64;

Query Match 13.4%; Score 85.5; DB 2; Length 691;
Best Local Similarity 27.1%; Pred. No. 10;
Matches 29; Conservative 14; Mismatches 51; Indels 13; Gaps 3;

QY 6 PHSGRVAVGYPAMGRCGRIGRWGQHGMNMLAIWHPKRY-----QSATTYQ 54
DB 241 PQRSLVLMALAPAMWRGRISAMATWQALAAVLFDPSAVLVGVGTWLSFGLVALIWA 300
DB 301 CSGRLHEGRQTAALRGQMAASVLSVLVLTGLYLPASPLISPLVNAVAI 347

QY 55 VDRSHDG-RTARVPDEITSTVS-GWLSLGTSQSLADELARAVRI 99
DB 301 CSGRLHEGRQTAALRGQMAASVLSVLVLTGLYLPASPLISPLVNAVAI 347

RESULT 6
Q9JVB8_NEIMA PRELIMINARY; PRT; 691 AA.
ID Q9JVB8_NEIMA PRELIMINARY; PRT; 691 AA.
AC Q9JVB8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Competence protein.
GN Name=comA; OrderedLocustNames=NMA0906;
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxId=56569;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=22491 / Serogroup A / Serotype 4A;
MEDLINE=20222556; PubMed=10761919; DOI=10.1038/3500665;
RA Parthill J., Achtman M., James K.D., Bentley S.D., Churcher C.M.,
RA Klee S.R., Morelli P., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N.,
RA Jørgen K., Leach S., Mout R., Mungall K.L., Quail M.A.,
RA Rajendram M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;

```

```

RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
DR EMBL; AL162754; CAB84183.1; -; Genomic_DNA.
DR PIR; B81937; B81937.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0030420; P:establishment of competence for transformation; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001279; Blactmase-like.
DR InterPro; IPR004477; ComEC_N-term.
DR InterPro; IPR004797; ComEC_Rec2.
DR Pfam; PF03772; Competence; 1.
DR Pfam; PF00753; Lactamase_B; 1.
DR TIGRFAMs; TIGR00360; ComEC_N-term; 1.
DR TIGRFAMs; TIGR00361; ComEC_Rec2; 1.
KW Complete proteome.
SQ SEQUENCE 691 AA; 74435 MW; D4924B8F446D8D9 CRC64;

Query Match 13.1%; Score 83.5; DB 2; Length 691;
Best Local Similarity 26.2%; Pred. No. 16;
Matches 28; Conservative 15; Mismatches 51; Indels 13; Gaps 3;

QY 6 PHSGRVAVGYPAMGRCGRIGRWGQHGMNMLAIWHPKRY-----QSATTYQ 54
DB 241 PQRSLVLMALAPAMWRGRISAMATWQALAAVLFDPSAVLVGVGTWLSFGLVALIWA 300
DB 301 CSGRLHEGRQTAALRGQMAASVLSVLVLTGLYLPASPLISPLVNAVAI 347

QY 55 VDRSHDG-RTARVPDEITSTVS-GWLSLGTSQSLADELARAVRI 99
DB 301 CSGRLHEGRQTAALRGQMAASVLSVLVLTGLYLPASPLISPLVNAVAI 347

RESULT 7
Q5YS68_NOCFA PRELIMINARY; PRT; 331 AA.
ID Q5YS68_NOCFA PRELIMINARY; PRT; 331 AA.
AC Q5YS68;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Putative dehydrogenase.
GN OrderedLocustNames=nfa41240;
OS Nocardia farcinica.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Nocardiaceae; Nocardia.
OX NCBI_TaxId=37329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IFM 10152;
RX PubMed=15466710; DOI=10.1073/pnas.0406410101;
RA Ishikawa J., Yamashita A., Mikami Y., Hoshino Y., Kurita H., Hotta K.,
RA Shida T., Hattori M.;
RT "The complete genomic sequence of Nocardia farcinica IFM 10152."
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
DE EMBL; AP006618; BAD58973.1; -; Genomic_DNA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR002085; ADH_SF_Zn.
DR Pfam; PF00107; ADH_zinc_N; 1.
KW Complete proteome.
SQ SEQUENCE 331 AA; 33227 MW; 8FE58428946EB29C CRC64;

Query Match 13.0%; Score 82.5; DB 2; Length 331;
Best Local Similarity 29.9%; Pred. No. 9;
Matches 38; Conservative 13; Mismatches 45; Indels 31; Gaps 6;

QY 10 GIRAVPYAMGRCGRIGRWGQHGMNMLAIWHPKRYQSATTYQV-----TDRS----H 60
DB 84 GARVVGPAPAGGIGGGVPTGVAERAL-----VTBQTLVAVPAGLIDRAVALVH 133
QY 61 DGRTRV-----PGEIST--VSGWLSLGTD---SPLADELARVIRIGDWPAAVAI 108
DB 134 DGRTLAAPDRATVQPSGLVLTVAAGGLGTLTOLAHRAQAKVVAARGAKELEAHL 193

```

```
Qy 109 GRLHSE 115
Db 194 GAHLSD 200

RESULT 8
Q7VVV9 BORPE PRELIMINARY; PRT; 281 AA.
ID Q7VVV9 BORPE PRELIMINARY; PRT; 281 AA.
AC Q7VVV9;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Putative secretory protein.
GN OrderedLocNames=BP2527;
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OK NCBI_TaxId=520;

NUCLEOTIDE SEQUENCE.
RC STRAIN-Tolama I / ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parthill J., Sebathia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdeno-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Felwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrett B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640418; CAE42801.1; -; Genomic_DNA.
DR InterPro; IPR010728; CPAB.
DR Pfam; PF06981; CPAB; 1.
KW Complete proteome.
SQ SEQUENCE 281 AA; 29193 MW; 506C7B9AB3406098 CRC64;

Query Match 12.9%; Score 82; DB 2; Length 281;
Best Local Similarity 30.6%; Pred. No. 8.5;
Matches 30; Conservative 11; Mismatches 29; Indels 28; Gaps 5;

Qy 10 GIRAVPYAMAGRCGRIGMGVHOEAMNMLAIWPRKYQSAITYQVTRSDH-GRTARVP 68
Db 9 GIRALGAYAFALVAGSLAAMARE-----HIQRVHDIEQARVP 48

69 GDEITSTVSGWLSLGTQSPPLADELARAVRGDMPAY 106
49 -MVASLVAAADLEAGT-PLHDEL---VAVDMFOAW 79

RESULT 9
Q7WKL0 BORBR PRELIMINARY; PRT; 277 AA.
ID Q7WKL0 BORBR PRELIMINARY; PRT; 277 AA.
AC Q7WKL0;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Putative exported protein.
GN OrderedLocNames=BB2096;
OS Bordetella bronchiseptica (Alcaligenes bronchiseptica).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OK NCBI_TaxId=518;

NUCLEOTIDE SEQUENCE.
RC STRAIN-RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parthill J., Sebathia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
```

```
RA Cerdeno-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Felwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrett B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640443; CAE32592.1; -; Genomic_DNA.
DR InterPro; IPR010728; CPAB.
DR Pfam; PF06981; CPAB; 1.
KW Complete proteome.
SQ SEQUENCE 277 AA; 28909 MW; B7731D268996BD62 CRC64;

Query Match 12.7%; Score 81; DB 2; Length 277;
Best Local Similarity 30.6%; Pred. No. 10;
Matches 30; Conservative 11; Mismatches 29; Indels 28; Gaps 5;

Qy 10 GIRAVPYAMAGRCGRIGMGVHOEAMNMLAIWPRKYQSAITYQVTRSDH-GRTARVP 68
Db 9 GIRALGAYAFALVAGSLAAMARE-----HIQRVHDIEQARVP 48

69 GDEITSTVSGWLSLGTQSPPLADELARAVRGDMPAY 106
49 -MVASLVAAADLEAGT-PLHDEL---VAVDMFOAW 79

RESULT 10
Q8ZT85 PYRAE PRELIMINARY; PRT; 263 AA.
ID Q8ZT85 PYRAE PRELIMINARY; PRT; 263 AA.
AC Q8ZT85;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Polyrenyl synthetase.
GN OrderedLocNames=PAE3294;
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC NCBI_TaxId=13773;

NUCLEOTIDE SEQUENCE.
RC STRAIN-IM2 / ATCC 51768 / DSM 7523;
RX MEDLINE=21664397; PubMed=11792869; DOI=10.1073/pnas.241636498;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AB009922; AAL64817.1; -; Genomic DNA.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000092; Polyrenyl_synth.
DR Pfam; PF00348; Polyrenyl_synth; 1.
DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
KW Complete proteome.
SQ SEQUENCE 263 AA; 29616 MW; AE265035F542FF67 CRC64;

Query Match 12.6%; Score 80.5; DB 2; Length 263;
Best Local Similarity 27.5%; Pred. No. 11;
Matches 25; Conservative 15; Mismatches 26; Indels 25; Gaps 4;

Qy 40 AIWPRKYQSAITYQVTRSDH-GRTARVP-----GDEITSTVSGWLSLGT 84
Db 62 ALMDPRLIQAATITIBLHVSLDDVDVMDKIEBRGKITPRVYVGGDDTILASDWL----- 117

85 TQSPPLADELARAVRGDMPAYA--IGEHLS 113
118 -----IASIRAVAVLGEVEVEYLVADVAQRLS 144
```





DR EMBL; BA000043; BAD77159.1; -, Genomic DNA.  
 DR GO; GO:0004177; F:aminopeptidase activity; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0006725; P:aromatic compound metabolism; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000073; A/b hydrolase.  
 DR InterPro; IPR001089; AB\_hydrolase.  
 DR InterPro; IPR000639; Epox\_hydrolase.  
 DR InterPro; IPR008262; Lipase\_AS.  
 DR InterPro; IPR002410; Peptidase\_S33.  
 DR InterPro; IPR000379; Ser esters.  
 DR Pfam; PF00561; Abhydrolase\_1; 1.  
 DR PRINTS; PRO0111; ABHYDROLASE.  
 DR PRINTS; PRO0412; EPOXYDOLASE.  
 DR PRINTS; PRO0793; PROAMNOPEPTASE.  
 DR PROSITE; PS00120; LIPASE\_SER; UNKNOWN\_1.  
 KM Amino-peptidase; Complete proteome; Hydrolase; Hypothetical protein.  
 SQ SEQUENCE 270 AA; 30278 MW; 9475BD73C480910 CRC64;

Query Match 12.4%; Score 79; DB 2; Length 270;  
 Best Local Similarity 23.6%; Pred. No. 16;  
 Matches 25; Conservative 16; Mismatches 41; Indels 24; Gaps 4;

OY 17 YMAAGRCGRIGRWGHOEAMM-----LAIWPKRQSATIYQVT--RSHDRT 64  
 DB 70 HAAALALALLDMEVGEQVAVVLCYSWGRLLALFAVWHPRVRRLVLESSSGLETEEEER 129  
 OY 65 ARVPGDE-----ITSTVSGW--LSELGTQSPDLAEARAV 98  
 DB 130 ARREADEALARKIETEGVAAPFDWDEKILPLFATQQLTPPVAAAR 175

RESULT 14  
 O33MF2 AMYMD  
 ID O33MF2 AMYMD PRELIMINARY; PRT; 163 AA.  
 AC O33MF2;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Amycolatopsis mediterranei (Nocardia mediterranei).  
 OC Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales;  
 OC Pseudonocardiales; Pseudonocardaceae; Amycolatopsids.  
 OX NCBI\_TaxId=33910;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=U32;  
 RX PubMed=15018097; DOI=10.1007/s00284-003-3997-5;  
 RA Wang W.W., Gao J., Zhao J., Zhao G., Jiang W.;  
 RT "Novel Two-Component System amrB-amkB Involved in the Regulation of  
 RT Central Carbohydrate Metabolism in Rifamycin SV-Producing  
 RT Amycolatopsis mediterranei U32.";  
 RL Curr. Microbiol. 48:14-19(2004).  
 DR EMBL; AJ318385; CAC42482.1; -, Genomic DNA.  
 KW Hypothetical protein.  
 FT NON TER 1  
 SQ SEQUENCE 163 AA; 18071 MW; 1546D37007E76ED1 CRC64;

Query Match 12.2%; Score 78; DB 2; Length 163;  
 Best Local Similarity 29.1%; Pred. No. 12;  
 Matches 25; Conservative 7; Mismatches 40; Indels 14; Gaps 2;

OY 6 PHSGRFVAVPYAMAGRCGRIGRWGHOEAMMMLAIWPKRQSATIYQVTRSHDGRYA 65  
 DB 4 PRGTG--ATRPVPRARHRRKLTGRKXDEANSLAMKVMHXSRYQ-----DDHYA 49  
 OY 66 RVPGDEITSTVSGWSELGTQSPDLAD 91  
 DB 50 DVAGTEQVWVTGQDWLSRFGLPPPAD 75

RESULT 15  
 O88E24\_PSEPK

ID O88E24\_PSEPK PRELIMINARY; PRT; 456 AA.  
 AC O88E24;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE DNA repair protein Rada.  
 GN Name=rada; OrderedLocustNames=PP4644;  
 OS Pseudomonas putida (strain KT2440).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxId=160488;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22423060; PubMed=12534463;  
 RX DOI=10.1046/j.1462-2920.2002.00366.x;  
 RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,  
 RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,  
 RA Brinkac L.M., Beanan M.J., DeBoy R.T., Daugherty S.C., Kolonay J.F.,  
 RA Madupu R., Nelson W.C., White O., Peterson J.D., Khoult H.M.,  
 RA Hance I., Chris Lee P., Holtzapfel E.K., Scanlan D., Tran K.,  
 RA Moazzez A., Utterback T.R., Rizzo M., Lee K., Kosack D., Moestl D.,  
 RA Medler H., Lauber J., Stjepandic D., Hohnesiel J., Straetz M., Helm S.,  
 RA Kiewitz C., Eisen J.A., Timmis K.N., Duesterhoeft A., Tuemmler B.,  
 RA Fraser C.M.;  
 RA "Complete genome sequence and comparative analysis of the  
 RT metabolically versatile Pseudomonas putida KT2440.";  
 RT Environ. Microbiol. 4:799-808(2002).  
 DR EMBL; AE016791; AA070217.1; -, Genomic DNA.  
 DR TIGR; PP4644; -.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004176; F:ATP-dependent peptidase activity; IEA.  
 DR GO; GO:0003684; F:damaged DNA binding; IEA.  
 DR GO; GO:0008094; F:DNA-dependent ATPase activity; IEA.  
 DR GO; GO:0017111; F:nucleoside-triphosphatase activity; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.  
 DR GO; GO:0006310; P:DNA recombination; IEA.  
 DR GO; GO:0006281; P:DNA repair; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR001984; Peptidase\_S16.  
 DR InterPro; IPR004504; Rada.  
 DR InterPro; IPR001553; Rech.  
 DR PRINTS; PRO0830; ENDOLAPTASE.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR00416; sma; 1.  
 DR PROSITE; PS0162; RECA\_2; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 456 AA; 48741 MW; 36CF6E54B5FSD265 CRC64;

Query Match 12.2%; Score 77.5; DB 2; Length 456;  
 Best Local Similarity 23.3%; Pred. No. 40;  
 Matches 30; Conservative 15; Mismatches 43; Indels 41; Gaps 6;

OY 19 WAGRCGRIGRWGHOEAMM-----NLAIWPKRQSATIYQVT----- 56  
 DB 20 WAGQCGEGGANNTLVETVIESGGAAPSSGAGWGTGQAQIKTLAEVSVBEI PRETTISST 79  
 OY 57 --DR-----SHDGRFARVPGDEITSTVSGWSELGTQSPDLAEARIGDWPAAYAI 109  
 DB 80 ELDRVLGGGLVDGVSIVLIGD-----PGIGKSTITLQTLCLN-IAVG-MPALVYTG 127  
 OY 110 EHLSEYAV 118  
 DB 128 EESQQQVAM 136

Search completed: March 23, 2006, 05:24:32  
 Job time : 31.6915 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using SW model

Run on: March 23, 2006, 05:24:53 ; Search time 8.1844 Seconds  
(without alignments)  
1212.194 Million cell updates/sec

Title: US-10-617-038-30

Perfect score: 637  
Sequence: 1 MSTOPRRHSGIRAVGPYAWA.....DWPAAVYAGHLSVEIAVAV 120

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5 COMB.pep.\*  
2: /cgn2\_6/prodata/1/1aa/6 COMB.pep.\*  
3: /cgn2\_6/prodata/1/1aa/7 COMB.pep.\*  
4: /cgn2\_6/prodata/1/1aa/8 COMB.pep.\*  
5: /cgn2\_6/prodata/1/1aa/9 COMB.pep.\*  
6: /cgn2\_6/prodata/1/1aa/Backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	12.2	310	2	US-09-252-991A-27948
2	75	11.8	510	2	US-09-252-991A-25258
3	74	11.6	310	2	US-09-252-991A-25258
4	73	11.5	449	2	US-10-104-047-2749
5	73	11.5	868	2	US-09-949-016-9731
6	71	11.1	292	1	US-08-737-526-4
7	71	11.1	292	2	US-09-098-580-4
8	70	11.0	2517	1	US-08-801-263A-5
9	70	11.0	2517	2	US-09-102-248-5
10	70	11.0	2517	2	US-09-167-764-5
11	69.5	10.9	329	2	US-09-252-991A-27332
12	69.5	10.9	630	2	US-09-252-991A-31264
13	69	10.8	625	2	US-09-489-039A-9779
14	68.5	10.8	451	2	US-09-477-962-118
15	68.5	10.8	544	2	US-09-976-594-699
16	68.5	10.8	544	2	US-09-919-039-270
17	68	10.7	571	2	US-10-104-047-3814
18	68	10.7	575	2	US-09-270-767-44868
19	67.5	10.6	343	2	US-09-252-991A-11876
20	67.5	10.6	414	2	US-09-252-991A-21182
21	67	10.5	522	2	US-09-949-016-9740
22	67	10.5	522	2	US-09-949-016-9740
23	66	10.4	201	2	US-09-252-991A-16881
24	66	10.4	209	2	US-09-252-991A-20905
25	66	10.4	346	2	US-09-252-991A-32864
26	66	10.4	535	2	US-09-252-991A-28410
27	66	10.4	591	2	US-09-561-077C-18

28	66	10.4	591	2	US-09-221-014-18	Sequence 18, Appl
29	66	10.4	718	2	US-09-252-991A-28229	Sequence 28229, A
30	65	10.2	152	2	US-09-252-991A-18469	Sequence 18469, A
31	65	10.2	359	2	US-09-252-991A-33080	Sequence 33080, A
32	65	10.2	513	2	US-09-270-767-46200	Sequence 46200, A
33	65	10.2	534	2	US-09-252-991A-30678	Sequence 30678, A
34	65	10.2	575	2	US-09-489-039A-8630	Sequence 8630, Ap
35	64.5	10.1	366	2	US-09-902-540-13525	Sequence 13525, A
36	64.5	10.1	565	2	US-09-252-991A-28571	Sequence 28571, A
37	64.5	10.1	1467	2	US-09-252-991A-17657	Sequence 17657, A
38	64	10.0	568	2	US-09-489-039A-9055	Sequence 9055, Ap
39	64	10.0	777	2	US-09-252-991A-31021	Sequence 31021, A
40	64	10.0	1958	2	US-10-152-886-93	Sequence 93, Appl
41	64	10.0	2446	2	US-09-949-016-11439	Sequence 11439, A
42	64	10.0	6095	2	US-09-144-085-2	Sequence 2, Appl1
43	63.5	10.0	377	2	US-09-679-279-3	Sequence 3, Appl1
44	63.5	10.0	423	2	US-09-902-540-9731	Sequence 9731, Ap
45	63.5	10.0	573	2	US-10-067-422-10	Sequence 10, Appl

#### ALIGNMENTS

```

RESULT 1
US-09-252-991A-27948
; Sequence 27948, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27948
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27948

Query Match      12.2%; Score 78; DB 2; Length 310;
Best Local Similarity 26.0%; Pred. No. 0.58;
Matches 27; Conservative 4; Mismatches 33; Indels 40; Gaps 3;

OY      9 SGIRAVGPYAMRGCRIGRMGCVHDEAMNMLIAIWHPRKVQSAITVQVTDHSHGRTARVP 68
Db      237 AGRRCAGKAGSAGAGRG-----DGGTARRA 263

OY      69 GDEITSVSGWLSLGTGSPPLADELARAVRIGDWPAAVAGHLL 112
Db      264 GERQAPAGTWRQRLG---QCROETARRLRIRGOWPDA---GGHLL 300

RESULT 2
US-09-252-991A-25258
; Sequence 25258, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27

```

NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 25258  
LENGTH: 510  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25258

Query Match 11.8%; Score 75; DB 2; Length 510;  
Best Local Similarity 26.3%; Pred. No. 2.6;  
Matches 31; Conservative 19; Mismatches 52; Indels 16; Gaps 4;

QY 11 IAAVPMYAGRCGRGNGHDEAMNLAIMHPRVGSATITVQVDRSHDRTARV--- 67  
DB 261 IEAAVPEAQSGLP-----QVGTQAVRLAAPEGEVLQARIYSILPEANRDRTRVARVLE 314  
QY 68 ---PGDEITSTVSGMLSELTGOSP-LADELARAVRIGDMPAAVYAI---GEHLSVEIAY 118  
DB 315 MANPGORLAKMSGGQIALKGNEDPALLVPESEAVIRIGKALAYVVDGPKFHPVEVQV 372

RESULT 3  
US-09-252-991A-26360  
Sequence 26360, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.116  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 26360  
LENGTH: 330  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26360

Query Match 11.6%; Score 74; DB 2; Length 330;  
Best Local Similarity 25.2%; Pred. No. 1.9;  
Matches 31; Conservative 13; Mismatches 59; Indels 20; Gaps 4;

QY 2 STQRRHSGIRAVGPYAA-----GRCGRIGRW-----GVHGEAMNLAIMH 43  
DB 26 TSRRRPGPELPMAAPVAMARRRLGSRSSRSRGWPSRVCTTSMPTPRAOSSAMIAIGA 85  
QY 44 PKVOSATITVQVDRS-HDGRTRAVPGDEITSTVSGMLSELTGOSPLADELARAVRIGDM 102  
DB 66 RRRRLTSIPMSAVPSPAQSRCSMTSSAVCSSCVRLTAQTPCTPR-RRCIARCG 144  
QY 103 PAA 105  
DB 145 PAA 147

RESULT 4  
US-10-104-047-2749  
Sequence 2749, Application US/10104047  
Patent No. 6943241  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: NO. 6943241el full length CDNA  
FILE REFERENCE: H1-A0105  
CURRENT APPLICATION NUMBER: US/10/104,047  
CURRENT FILING DATE: 2002-03-25  
PRIOR APPLICATION NUMBER:  
PRIOR FILING DATE:  
NUMBER OF SEQ ID NOS: 4096  
SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2749  
LENGTH: 449  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-104-047-2749

Query Match 11.5%; Score 73; DB 2; Length 449;  
Best Local Similarity 28.1%; Pred. No. 3.8;  
Matches 18; Conservative 15; Mismatches 27; Indels 4; Gaps 1;

QY 54 QVDRSHDGRTRAVPGDEITSTVSGMLSELTGOSPLADELARAVRIGDMPAAVIGELHS 113  
DB 249 ELSDVSCGGSXATTP---ASTANSDVATITPTDPLKEENGFPVATDAPNKSEISKIE 304  
QY 114 VEIA 117  
DB 305 VQVA 308

RESULT 5  
US-09-949-016-9731  
Sequence 9731, Application US/09949016  
Patent No. 6012339

GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9731  
LENGTH: 868  
TYPE: PRT  
ORGANISM: Human  
US-09-949-016-9731

Query Match 11.5%; Score 73; DB 2; Length 868;  
Best Local Similarity 28.1%; Pred. No. 9.1;  
Matches 18; Conservative 15; Mismatches 27; Indels 4; Gaps 1;

QY 54 QVDRSHDGRTRAVPGDEITSTVSGMLSELTGOSPLADELARAVRIGDMPAAVIGELHS 113  
DB 186 ELSDVSCGGSXATTP---ASTANSDVATITPTDPLKEENGFPVATDAPNKSEISKIE 241  
QY 114 VEIA 117  
DB 242 VQVA 245

RESULT 6  
US-08-737-526-4  
Sequence 4, Application US/08737526  
Patent No. 5871966  
GENERAL INFORMATION:  
APPLICANT: Kofoed, Lene Venke  
APPLICANT: Andersen, Lene No. 5871966boe  
APPLICANT: Kauppinen, Markus Sakari  
APPLICANT: Christgau, Stephan  
TITLE OF INVENTION: An Enzyme With Endo-1,3(4)-B-Glucanase  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESS: No. 5871966o No. 5871966disk of No. 5871966th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York

STATE: NY  
COUNTRY: USA  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/737,526  
FILING DATE: 08-NOV-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Valecia, Gregg A  
REGISTRATION NUMBER: 35,127  
REFERENCE/DOCKET NUMBER: 4174.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 292 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-737-526-4

Query Match 11.1%; Score 71; DB 1; Length 292;  
Best Local Similarity 26.5%; Pred. No. 3.7;  
Matches 35; Conservative 18; Mismatches 51; Indels 28; Gaps 8;

QY 4 QPRHSGIRAVGPYVMAAGRCG---RIGRWGVHQAEMMLAIWPKVKOSATTIYQVTDSSH 60  
21 QPPTYSGFNLVWTDTPFAGNGGTSPNQNNMI---ITGNLNVAAEQETYSSTANV--QLS 75  
DB 61 DGRTR-VP---GDEITSTVSGM---LSELGTOSPLADELAR---AVRIGD----- 101  
76 GGSITQLVPMRDSKGTSTFGMTSGRLSKYTFPPAAGKVTRELAIRFGSMAQNAKOG 135  
QY 102 -WPAAYAIGEHL 112  
DB 136 IWPFAFMLGDSL 147

RESULT 7  
US-09-098-580-4  
Sequence 4, Application US/09098580  
Patent No. 6140096  
GENERAL INFORMATION:  
APPLICANT: Kofod, Lene Venke  
APPLICANT: Andersen, Lene No. 6140096boe  
APPLICANT: Kauppinen, Markus Sakari  
APPLICANT: Christgau, Stephan  
TITLE OF INVENTION: An Enzyme with Endo-1,3(4)-B-Glucanase  
TITLE OF INVENTION: Activity  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSES: No. 61400960 No. 6140096disk of No. 6140096th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/098,580  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/737,526  
FILING DATE: 08-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Valecia, Gregg A  
REGISTRATION NUMBER: 35,127  
REFERENCE/DOCKET NUMBER: 4174.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 292 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-098-580-4

Query Match 11.1%; Score 71; DB 2; Length 292;  
Best Local Similarity 26.5%; Pred. No. 3.7;  
Matches 35; Conservative 18; Mismatches 51; Indels 28; Gaps 8;

QY 4 QPRHSGIRAVGPYVMAAGRCG---RIGRWGVHQAEMMLAIWPKVKOSATTIYQVTDSSH 60  
21 QPPTYSGFNLVWTDTPFAGNGGTSPNQNNMI---ITGNLNVAAEQETYSSTANV--QLS 75  
DB 61 DGRTR-VP---GDEITSTVSGM---LSELGTOSPLADELAR---AVRIGD----- 101  
76 GGSITQLVPMRDSKGTSTFGMTSGRLSKYTFPPAAGKVTRELAIRFGSMAQNAKOG 135  
QY 102 -WPAAYAIGEHL 112  
DB 136 IWPFAFMLGDSL 147

RESULT 8  
US-08-801-263A-5  
Sequence 5, Application US/08801263A  
Patent No. 5811407  
GENERAL INFORMATION:  
APPLICANT: Johnston, Robert E.  
APPLICANT: Davis, Nancy L.  
APPLICANT: Simpson, Dennis A.  
TITLE OF INVENTION: System for the in Vivo Delivery and  
TITLE OF INVENTION: Expression of Heterologous Genes in the Bone Marrow  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Bell Seltzer Park & Gibson, P.A.  
STREET: 1211 East Morehead Street  
CITY: Charlotte  
STATE: No. 5811407th Carolina  
COUNTRY: USA  
ZIP: 28234  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/801,263A  
FILING DATE: 19-FEB-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Sibley, Kenneth D.  
REGISTRATION NUMBER: 31,665  
REFERENCE/DOCKET NUMBER: 5470-147  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-881-3175  
TELEFAX: 919-881-3175  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2517 amino acids



```
RESULT 11
US-09-252-991A-27332
; Sequence 27332, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27332
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27332

Query Match
Best Local Similarity 10.9%; Score 69.5; DB 2; Length 329;
Matches 35; Conservative 12; Mismatches 55; Indels 13; Gaps 5;

QY 4 QRRHSGIRAVGPFYANAGRCGRIGRWGVQEMAMNLAIW-HPRKVQSAITYQVTDNRSHD 61
DB 216 QRRPAGRG-RALVPRARG-FARLRLSGAGQGRPARVWRGRPRRCRAGACVSRARG 273

QY 62 GRT-----ARVPGDEITSTVSGWLSLGTQSPLELARAVERIGDWPAAVAIGE 110
DB 274 QRTDTRACRRPRADRLSSAVGALS---GEPRIAGQAPVAGGQGPQAGHE 325

RESULT 12
US-09-252-991A-31264
; Sequence 31264, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31264
; LENGTH: 630
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31264

Query Match
Best Local Similarity 10.9%; Score 69.5; DB 2; Length 630;
Matches 22; Conservative 14; Mismatches 35; Indels 7; Gaps 3;

QY 43 HPRKVSATITYQVTDNRSHD-DGRTARVPDGRITSTVSGWLSLGTQSPLELARAVERI 99
DB 165 HWPSSSATELSGARPARHAQSKTLPIDPKLAEIFRAVSRLL---ERADETLAAENV 240

QY 100 GDMPAAYALGEHLISVEIA 117
DB 241 SDMQAQSYSGTYNATIA 258

RESULT 13
US-09-489-039A-9779
; Sequence 9779, Application US/09489039A
```

```
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9779
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9779

Query Match
Best Local Similarity 10.8%; Score 69; DB 2; Length 625;
Matches 22; Conservative 11; Mismatches 33; Indels 30; Gaps 3;

QY 28 RWGVQEMAMNL-----ALMHPRKVSATITYQVTDNRSHDGRARVPGDEITSTV 76
DB 443 QMGKAVSLMEIPTTGRTIDNIVCFWQPEKAVKA-----GDELDPRY 484

QY 77 S-GWLSLGTQSPLELARAVERIGDWPAAVAIGE 111
DB 485 RLYWSAQPVSTPLARVLATRTGCGFPEGMAFGSH 520

RESULT 14
US-09-477-962-118
; Sequence 118, Application US/09477962
; Patent No. 6927286
; GENERAL INFORMATION:
; APPLICANT: SHEN, BEN
; APPLICANT: DU, LIANGCHENG
; APPLICANT: SANCHEZ, CESAR
; APPLICANT: CHEN, MEI
; APPLICANT: EDWARDS, DANIEL J.
; TITLE OF INVENTION: BLEBOMYCIN GENE CLUSTER COMPONENTS AND THEIR USES
; FILE REFERENCE: 407T-895820US
; CURRENT APPLICATION NUMBER: US/09/477,962
; PRIOR FILING DATE: 2000-01-05
; PRIOR APPLICATION NUMBER: 60/115,435
; PRIOR FILING DATE: 1999-01-06
; PRIOR APPLICATION NUMBER: 60/118,848
; PRIOR FILING DATE: 1999-02-05
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 118
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Streptomyces verticillius
; FEATURE:
; OTHER INFORMATION: ORF33
US-09-477-962-118

Query Match
Best Local Similarity 10.8%; Score 68.5; DB 2; Length 451;
Matches 30; Conservative 13; Mismatches 46; Indels 19; Gaps 4;

QY 6 PRRSGIRAVGPFYANAGRCGRIGRWGVQEMAMNLAIWHPRKVSATITYQVTDNRSHDGR 65
DB 58 PRRAG-----LGRVHSSQAALVAL-GPMVAATTTGVAAGDGHQARG 101

QY 66 RVPDGRITSTVSGWLSLGTQSPLELARAVERIGDWPAAVAIGEHL 112
DB 102 QVTGVGVREVAAGGDDDELGAEDRPHARG-RUDDGLGMAAERLADLL 147

RESULT 15
US-09-976-594-699
```

```
; Sequence 699, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 699
; LENGTH: 544
; TYPE: PRF
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Inocyte ID No. 6673549 4113161CD1
US-09-976-594-699
```

```
Query Match 10.8%; Score 68.5; DB 2; Length 544;
Best Local Similarity 24.8%; Pred. NO. 17;
Matches 37; Conservative 16; Mismatches 43; Indels 53; Gaps 9;

QY 10 GIRAVGPYAMGRGCRIGRW-----GVHDEAMMNLAIWHPKVVQSATIIYQVTD 58
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25 GALGPGPHRW-GPCG--GSMWQKFYQDGGRGGLGSEDIRRAREARPRKTPRP---QLSDR 78
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 59 SHDGRARVPGEITSTVS-----GWLSELTQ-----SPLAD 91
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 79 S---RRKKVPASGISRLANFGGLAVGLGLVLAEMAKSMGGRLOSGSGGIDSPFLS 135
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 92 ELARAVRIGDW-----PAAYAIGEHLSVE 115
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 136 E-AMMERIVQTLCTVRGALKKVGQMLSTQ 163
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: March 23, 2006, 05:30:20  
Job time : 9.2844 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2006, 06:14:33 ; Search time 25.2352 Seconds  
(without alignments)  
1966.885 Million cell updates/sec

Title: US-10-617-038-30

Perfect score: 637  
Sequence: 1 MSTOPRRSGIRAVGPYMA.....DMPAAVAIGEHLSVEIAVAV 120

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA\_Main:\*  
1: /cgn2\_6/prodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/prodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/prodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/prodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
5: /cgn2\_6/prodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
6: /cgn2\_6/prodata/1/pubppaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	637	100.0	120	4 US-10-617-038-30	Sequence 30, Appl
2	80	12.6	551	4 US-10-425-114-39387	Sequence 39387, A
3	79.5	12.5	1167	5 US-10-794-514A-328	Sequence 328, App
4	77.5	12.2	456	4 US-10-282-122A-68099	Sequence 68099, A
5	77	12.1	278	4 US-10-424-599-217918	Sequence 217918,
6	75.5	11.9	272	4 US-10-437-963-170212	Sequence 170212,
7	75	11.8	8605	6 US-11-097-143-28128	Sequence 28128, A
8	74.5	11.7	279	4 US-10-437-963-145941	Sequence 145941,
9	74	11.6	284	4 US-10-156-761-8113	Sequence 8113, Ap
10	74	11.6	535	4 US-10-282-122A-62103	Sequence 62103, A
11	73	11.5	69	3 US-09-864-761-43789	Sequence 43789, A
12	73	11.5	258	4 US-10-106-698-6371	Sequence 6371, Ap
13	73	11.5	365	4 US-10-732-923-10308	Sequence 10308, A
14	73	11.5	449	4 US-10-104-047-2749	Sequence 2749, Ap
15	73	11.5	715	3 US-09-815-242-10450	Sequence 10450, A
16	73	11.5	715	4 US-10-369-493-23660	Sequence 23660, A
17	73	11.5	715	4 US-10-282-122A-43223	Sequence 43223, A
18	73	11.5	715	4 US-10-282-122A-75843	Sequence 75843, A
19	73	11.5	766	4 US-10-205-219-157	Sequence 157, App
20	73	11.5	1277	4 US-10-276-774-2154	Sequence 2154, Ap
21	72.5	11.4	210	4 US-10-437-963-147829	Sequence 147829,
22	72.5	11.4	492	5 US-10-450-763-45094	Sequence 45094, A
23	72.5	11.4	492	4 US-10-156-761-10210	Sequence 10210, A
24	71	11.1	472	5 US-10-732-923-1072	Sequence 1072, Ap
25	71	11.1	19723	4 US-10-084-846A-5	Sequence 5, Appl
26	70.5	11.1	99	4 US-10-437-963-159656	Sequence 159656,
27	70.5	11.1	281	4 US-10-369-493-4003	Sequence 4003, Ap

28	70.5	11.1	753	4 US-10-282-122A-49874	Sequence 49874, A
29	70.5	11.1	954	4 US-10-437-963-184007	Sequence 184007, A
30	70	11.0	838	5 US-10-779-597-95	Sequence 95, Appl
31	70	11.0	3362	4 US-10-376-083-6	Sequence 6, Appl
32	69.5	10.9	235	5 US-10-481-265-19	Sequence 19, Appl
33	69.5	10.9	330	4 US-10-424-599-153232	Sequence 153232,
34	69.5	10.9	453	4 US-10-156-761-1136	Sequence 14136, A
35	69.5	10.9	696	3 US-09-738-628-5259	Sequence 5259, Ap
36	69.5	10.9	981	4 US-10-156-761-15071	Sequence 15071, A
37	69	10.8	327	4 US-10-437-963-201707	Sequence 201707,
38	69	10.8	391	4 US-10-156-761-9791	Sequence 9791, Ap
39	69	10.8	476	4 US-10-425-114-52031	Sequence 52031, A
40	69	10.8	646	4 US-10-437-963-136807	Sequence 136807,
41	69	10.8	921	4 US-10-467-685-14	Sequence 14, Appl
42	69	10.8	927	4 US-10-408-765A-2938	Sequence 2938, Ap
43	69	10.8	1380	5 US-10-450-763-33073	Sequence 33073, A
44	68.5	10.8	238	4 US-10-425-115-246803	Sequence 246803,
45	68.5	10.8	365	5 US-10-732-923-10307	Sequence 10307, A

## ALIGNMENTS

```
RESULT 1
US-10-617-038-30
; Sequence 30, Application US/10617038
; Publication No. US20040057963A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Peter
; APPLICANT: Rosenkrands, Ida
; APPLICANT: Steyn, Anette
; TITLE OF INVENTION: Therapeutic TB Vaccine
; FILE REFERENCE: SSI5AUSA
; CURRENT APPLICATION NUMBER: US/10/617,038
; PRIOR FILING DATE: 2003-07-11
; PRIOR APPLICATION NUMBER: DK PA 2002 01098
; PRIOR FILING DATE: 2002-07-13
; PRIOR APPLICATION NUMBER: US 60/401,725
; PRIOR FILING DATE: 2002-08-07
; NUMBER OF SEQ ID NOS: 187
; SOFTWARE: PatentIn version 3.1.2
; SEQ ID NO 30
; LENGTH: 120
; TYPE: PRP
; ORGANISM: Mycobacterium tuberculosis
US-10-617-038-30

Query Match      100.0%; Score 637; DB 4; Length 120;
Best Local Similarity 100.0%; Pred. No. 1.4e-63;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MSTOPRRSGIRAVGPYMAAGRCGRIGRWGQEAAMNLAIWPPKVOATITYOTDRSH 60
      |||
DB      1 MSTOPRRSGIRAVGPYMAAGRCGRIGRWGQEAAMNLAIWPPKVOATITYOTDRSH 60
      |||

QY      61 DGRTRVPEDEITTSVSGWLSLGTQSPDLADLAVRIGDMPAAVAIGEHLSVEIAVAV 120
      |||
DB      61 DGRTRVPEDEITTSVSGWLSLGTQSPDLADLAVRIGDMPAAVAIGEHLSVEIAVAV 120
      |||

RESULT 2
US-10-425-114-39387
; Sequence 39387, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
```





```

Oy      4 QRPRLSGIRAVG---PYA-----WAGCGRIGVGHQAMMLIAIWHPRKVSATIYQV 55
      :|||      |||      :|||      :|||      :
Db      71 RRPRLRGRRLEAARPRPYGRRRRPWAAHSSGREGGKXETRRRSIA--HPR--STATXEKA 122

```

AFFIDAVIT: YES, Filing  
 TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 FILE REFERENCE: 38-21(53221)B  
 CURRENT APPLICATION NUMBER: US/10/437,963  
 CURRENT FILING DATE: 2003-05-14

```
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 149941
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(279)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_50224C.1.pap
; US-10-437-963-149941

Query Match      11.7%; Score 74.5; DB 4; Length 279;
Best Local Similarity 29.0%; Pred. No. 8.7;
Matches 40; Conservative 13; Mismatches 52; Indels 33; Gaps 8;

QY      3  TORPRHSGIRAVGPVAMAGRCGR--IGRWGHOEAMNMLAIWHPKVSATITYQVTDNRSH 60
      120  TORPRRGDEARPPPTAVTGGACGCGDGRKAAAHGRSSGVAGARERTKAT-----TTRRE 175
      61  DERTA-----RVGDEI-----TSTVGWMLSELGTQSLADLBAVRIGDW-PAAY- 106
      176  EGRTSQGDYXRRGGSEELGMEKATATTYKVGWRRRRGMDSSP-----AREGDVFPVGR 227
      107  ---AIGE--HLSVEIYVA 119
      228  XREAVAMAHMTAAELAA 245
      DB

RESULT 9
US-10-156-761-8113
; Sequence 8113, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8113
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-8113

Query Match      11.6%; Score 74; DB 4; Length 284;
Best Local Similarity 25.0%; Pred. No. 10;
Matches 17; Conservative 12; Mismatches 35; Indels 4; Gaps 1;

QY      36  MNMLAIWHPKVSATITYQVTDNRSHGRTARVPGDEITSTVGWMLSELGTQSLADLAR 95
      88  VVNAGSLHAGLAAARELIARHGHDLKRVSHITGSDVFGSLDS----LQAGHCLPHLS 143
      DB

QY      96  AVRIGDW 103
      144  GQPLSDWP 151
      DB

RESULT 10
US-10-282-122A-62103
; Sequence 62103, Application US/10282122A
```

```
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyn, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.03A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-09-06
; PRIOR FILING DATE: 2000-09-06
; PRIOR FILING DATE: 2000-09-09
; PRIOR FILING DATE: 2000-09-09
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62103
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Mycobacterium avium
; US-10-282-122A-62103

Query Match      11.6%; Score 74; DB 4; Length 535;
Best Local Similarity 28.2%; Pred. No. 22;
Matches 35; Conservative 15; Mismatches 46; Indels 28; Gaps 7;

QY      10  GIRAVGPYAMAGRCGRIGRWGHOEAMNMLAIWHPKVSATITYQVTDNRSHGRTARVPG 69
      379  GIRVGFRTWPG-FARQDQIGTAFSTVMDIA-----FTLELACTSHPGTAYR--G 426
      DB

QY      70  DEIT-----STVGWMLSELGTQSLAD-----EL--ARAVRIGDW-----PAAYAIGEHLS 113
      427  REVAPMRGSLVGYLSGAADTVHDADTETGWELEFRRARIQGNWKAHLIPAVYGPQSQWL 486
      114  VEIA 117
      DB

QY      487  YDLA 490
      DB

RESULT 11
US-09-864-43789
; Sequence 43789, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
```

```

1  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
2  TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
3  FILE REFERENCE: Aecm1ca-X-1
4  CURRENT FILING DATE: 2001-05-23
5  PRIOR APPLICATION NUMBER: US 60/180,312
6  PRIOR FILING DATE: 2000-02-04
7  PRIOR APPLICATION NUMBER: US 60/207,456
8  PRIOR FILING DATE: 2000-05-26
9  PRIOR APPLICATION NUMBER: US 09/632,366
10 PRIOR FILING DATE: 2000-08-03
11 PRIOR APPLICATION NUMBER: GB 24263.6
12 PRIOR FILING DATE: 2000-10-04
13 PRIOR APPLICATION NUMBER: US 60/236,359
14 PRIOR FILING DATE: 2000-09-27
15 PRIOR APPLICATION NUMBER: PCT/US01/00666
16 PRIOR FILING DATE: 2001-01-30
17 PRIOR APPLICATION NUMBER: PCT/US01/00667
18 PRIOR FILING DATE: 2001-01-30
19 PRIOR APPLICATION NUMBER: PCT/US01/00664
20 PRIOR FILING DATE: 2001-01-30
21 PRIOR APPLICATION NUMBER: PCT/US01/00669
22 PRIOR FILING DATE: 2001-01-30
23 PRIOR APPLICATION NUMBER: PCT/US01/00665
24 PRIOR FILING DATE: 2001-01-30
25 PRIOR APPLICATION NUMBER: PCT/US01/00668
26 PRIOR FILING DATE: 2001-01-30
27 PRIOR APPLICATION NUMBER: PCT/US01/00663
28 PRIOR FILING DATE: 2001-01-30
29 PRIOR APPLICATION NUMBER: PCT/US01/00662
30 PRIOR FILING DATE: 2001-01-30
31 PRIOR APPLICATION NUMBER: PCT/US01/00661
32 PRIOR FILING DATE: 2001-01-30
33 PRIOR APPLICATION NUMBER: PCT/US01/00670
34 PRIOR FILING DATE: 2001-01-30
35 PRIOR APPLICATION NUMBER: US 60/234,687
36 PRIOR FILING DATE: 2000-09-21
37 PRIOR APPLICATION NUMBER: US 09/608,408
38 PRIOR FILING DATE: 2000-06-30
39 PRIOR APPLICATION NUMBER: US 09/774,203
40 PRIOR FILING DATE: 2001-01-29
41 NUMBER OF SEQ ID NOS: 49117
42 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
43 SEQ ID NO 43789
44 LENGTH: 69
45 TYPE: PRT
46 ORGANISM: Homo sapiens
47 FEATURE:
48 OTHER INFORMATION: MAP TO AC005920.1
49 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.68
50 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.68
51 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.53
52 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.69
53 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.61
54 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.97
55 OTHER INFORMATION: SWISSPROT HIT: P17691, EVALU8 8.80e-01
56 OTHER INFORMATION: EST_HUMAN HIT: A0143722.1, EVALU8 6.00e-25
57 US-09-864-761-43789

```

```

US-10-106-698-6371
? Sequence 6371, Application US/10106698
? Publication No. US20030109690A1
? GENERAL INFORMATION:
? APPLICANT: Ruben et al.
? TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
? FILE REFERENCE: PA005P1
? CURRENT APPLICATION NUMBER: US/10/106,698
? CURRENT FILING DATE: 2002-03-27
? PRIOR APPLICATION NUMBER: PCT/US00/26524
? PRIOR FILING DATE: 2000-09-28
? PRIOR APPLICATION NUMBER: US 60/157,137
? PRIOR FILING DATE: 1999-09-29
? PRIOR APPLICATION NUMBER: US 60/163,280
? PRIOR FILING DATE: 1999-11-03
? NUMBER OF SEQ ID NOS: 8564
? SOFTWARE: PatentIn Ver. 3.0
? SEQ ID NO 6371
? LENGTH: 258
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: MISC FEATURE
? LOCATION: (140)
? OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-6371

```

Query Match	11.5%	Score 73	DB 4	Length 258
Best Local Similarity	27.0%	Pred. No. 12		
Matches	37	Conservative	14	Mismatches 40
				Indels 46
				Gaps 9

  

QY	9	SGIRAVGPY--AWAGRCGRIG-----	RMGVHOEAMNLAIW-HPRKY	47
Db	121	STVREVGHSRSTGNCAXLAALADERGVAQINTEADAKRGIDREAL---	VWHSRKG	176
QY	48	QSATTYVTDRBSHDGRTARVPGDEITTSVSCML--SLGTQ--SPLADE-----	LARA	96
Db	177	KIITRAVVSDBRPNG-----AIYMTQWIGACNELVTENSPITKTPEYKCAVRV	228	
QY	97	VRIGDWPA--YAIGH	111	
Db	229	EPIADQRAAEQYVIDEY	245	

```

RESULT 13
US-10-732-923-10308
: Sequence 10308, Application US/10732923
: Publication No. US20050108791A1
: GENERAL INFORMATION:
: APPLICANT: Edgerton, Michael D
: TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
: FILE REFERENCE: 38-15 (55796)C
: CURRENT APPLICATION NUMBER: US/10/732,923
: CURRENT FILING DATE: 2003-12-10
: PRIOR APPLICATION NUMBER: 10/310,154
: PRIOR FILING DATE: 2002-12-04
: NUMBER OF SEQ. ID NOS: 24149
: SEQ. ID NO 10308
: LENGTH: 365
: TYPE: PR1
: ORGANISM: Escherichia coli CFT073
: US-10-732-923-10308

```

```

Query Match 11.5%; Score 73; DB 3; Length 69;
Best Local Similarity 28.1%; Pred. No. 2, 4;
Matches 18; Conservative 15; Mismatches 27; Indels 4; Gaps 1

Oy 54 QVTRBSHQRTARVPGDEITSTVSGMSELGTQSPPLADELARVARIGDMPAAVIAIGELHS 113
Db 2 ELSDVSQSGSKATTP---ASTANSDVARIPTDTPUKENEGFVKVTDAPNKSISKITE 57
Oy 114 VEIA 117
Db 58 VQVA 61

RESULT 12

```

```

Query Match      11.5% Score 73; DB 5; Length 365;
Best Local Similarity 26.0%; Pred. No. 18;
Matches 27; Conservative 17; Mismatches 32; Indels 28; Gaps 6

QY    19 WAGRCG-----RIGW-----GVHQE-AMNNLAIWMHPKVASATITYVTD RSHDGR TARV 67
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     69 YGAPGPIHSEQDIAMWKKTITAGVAHENGMHVMQLMHTGRISHASLQ----- 114
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

QY    68 PDEDTSTVSGWLSELGTOSPLADELAARAVI-GDWPRALALGE 110
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

Db 115 PGGC--APVAPSAISAGTSTSLRDENGCAIRVETSMPLALEIGE 156

RESULT 14  
US-10-104-047-2749  
; Sequence 2749, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20030236392A1 full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; PRIOR FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2749  
; LENGTH: 449  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-104-047-2749

Query Match 11.5%; Score 73; DB 4; Length 449;  
Best Local Similarity 28.1%; Pred. No. 23; Indels 4; Gaps 1;  
Matches 18; Conservative 15; Mismatches 27

QY 54 QVTRSHDGRTPRVGDEITSTVSGWSELGTQSPDLADLARAVRIGDMPAAVAIGEHLS 113  
DB 249 ELSDVSGGSKATTE---ASTANSDVATITPTDTLKEENEGFVATVDAPNKSEISKHIE 304  
QY 114 VEIA 117  
DB 305 VQVA 308

RESULT 15  
US-09-815-242-10450  
; Sequence 10450, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zysek, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-23  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10450  
; LENGTH: 715  
; TYPE: PRT

; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(715)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-815-242-10450

Query Match 11.5%; Score 73; DB 3; Length 715;  
Best Local Similarity 27.0%; Pred. No. 40;  
Matches 37; Conservative 14; Mismatches 40; Indels 46; Gaps 9;

QY 9 SGIRAVGPY---AWAGRCGRIG-----RWGVHQAAMMNLAIW-HPRKV 47  
DB 578 STVEVGHYSCRSMTGNCAALALADEGVAQINTEDAKRIGIEDAL---VYHSRKG 633  
QY 48 QSATIIYQVTRSHDGRTPRVGDEITSTVSGWL--SELGTQ--SPLADE-----LARA 96  
DB 634 KITRAQVSDRPNGK-----AIYMTYQWIGACNELVTENLSPITKTPYKCAVRV 685  
QY 97 VRIGDMPAA--YAIGEH 111  
DB 686 EPIADQRAAEQYVIDEX 702

Search completed: March 23, 2006, 06:28:20  
Job time : 26.2352 secs



```

; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARA, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2749
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-2749

```

Query Match 11.5%; Score 73; DB 7; Length 449;

Best Local Similarity 28.1%; Pred. No. 1.9; Mismatches 27; Indels 4; Gaps 1;

```

Matches 18; Conservative 15; Mismatches 27; Indels 4; Gaps 1;
QY 54 QVTRSHDGRTPARVPGDEITSTVSGWLSLGTQSPDLDELARAVRIGDMPAAVIGEHLS 113
DB 249 ELSDVSGGSKATTP---ASTANSDVATIPDTPLKEENEGFVKVTDAPNKSEISKHLE 304
QY 114 VEIA 117
DB 305 VQVA 308

```

## RESULT 3

```

US-10-467-657-2722
; Sequence 2722; Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: PIZZA Mariagrazia Rita
; APPLICANT: MASIIGNANT Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqMin99, version 1.04
; SEQ ID NO 2722
; LENGTH: 759
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-2722

```

Query Match 11.2%; Score 71.5; DB 6; Length 759;

Best Local Similarity 25.2%; Pred. No. 5.2; Mismatches 53; Indels 13; Gaps 3;

```

Matches 27; Conservative 14; Mismatches 53; Indels 13; Gaps 3;
QY 6 PRHSGRVAVPYAMGRCGRIGRWGVOEAMMNLAIWHPKRY-----QSATTYQ 54
DB 309 PQRSLVLMALAPAMWRRLSNAMATWQALAVLLFDPLAVLGVTMLSPGLVAAALITVA 368
QY 55 VTRSHDG-RTARVPDEITSTVS-GWLSLGTQSPDLDELARAVRI 99
DB 369 CAGCTYEGKQTPAVNGWMAAVSLVILGYLPLASLPLVPLVNAVSI 415

```

## RESULT 4

US-10-509-464-5

```

; Sequence 5; Application US/10509464
; Publication No. US20050244826A1
; GENERAL INFORMATION:
; APPLICANT: NIEHRS, CHRISTOP
; APPLICANT: MAO, BINGYU
; TITLE OF INVENTION: ASSOCIATIONS FOR DIAGNOSIS AND THERAPY OF DISEASES
; TITLE OF INVENTION: ASSOCIATED WITH ABERRANT EXPRESSION OF KEMEN AND/OR
; FILE REFERENCE: WNT
; CURRENT APPLICATION NUMBER: US/10/509,464
; CURRENT FILING DATE: 2004-09-27
; PRIOR APPLICATION NUMBER: PCT/EP03/03277
; PRIOR FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: EP 02 008 650.0
; PRIOR FILING DATE: 2002-04-17
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 3.3
; SEQ ID NO 5
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-509-464-5

```

Query Match 10.9%; Score 69.5; DB 6; Length 473;

Best Local Similarity 27.8%; Pred. No. 5; Mismatches 44; Indels 21; Gaps 4;

```

Matches 30; Conservative 13; Mismatches 44; Indels 21; Gaps 4;
QY 3 TORPHSGIRAVGPYAMGRCG---RIGRWGVOEAMMNLAIWHPKRYQSATTYQVTR 58
DB 361 TPSPHPQTPAPGSHSMAPSVGANSHRVEGTVV--GLATLLILVTVAVKILLHTVFX 418
QY 59 SH-----DGRTPARVPGD-----EITSTVSGWLSLGTQSPDL 91
DB 419 SHRVPASGLRDCRPGASGDITWTFYBPSTTISIFKKLKGQSQODD 466

```

## RESULT 5

```

US-11-146-428-47
; Sequence 47; Application US/11146428
; Publication No. US20060051847A1
; GENERAL INFORMATION:
; APPLICANT: GUNNARSSON, NINA KATARINA
; APPLICANT: FORSTER, JOCHEN
; APPLICANT: NEILSEN, JENS BREDEL
; TITLE OF INVENTION: METABOLICALLY ENGINEERED CELLS FOR THE
; TITLE OF INVENTION: PRODUCTION OF POLYUNSATURATED FATTY ACIDS
; FILE REFERENCE: 671306-2001.1
; CURRENT APPLICATION NUMBER: US/11/146,428
; CURRENT FILING DATE: 2005-06-06
; PRIOR APPLICATION NUMBER: 60/577,245
; PRIOR FILING DATE: 2004-06-04
; NUMBER OF SEQ ID NOS: 227
; SOFTWARE: Patent In Ver. 3.3
; SEQ ID NO 47
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Cryptococcus curvatus
US-11-146-428-47

```

Query Match 10.8%; Score 68.5; DB 7; Length 446;

Best Local Similarity 21.2%; Pred. No. 6; Mismatches 44; Indels 29; Gaps 4;

```

Matches 24; Conservative 16; Mismatches 44; Indels 29; Gaps 4;
QY 20 AGRCGRIGRWGVOE---EAMMNLAIW-----HPRKVQSATTI----- 52
DB 97 AHTAKILMFTTYQVLAVGVTGIVNVAHEGCHQAYSSKTIINNAVGVVHLSILVPHYS 156
QY 53 QVTRSHDGRTPARVPGDEITSTVSGWLSLGTQSPDLDELARAVRIGDMPAA 105
DB 157 WRISHGRHATVGHLTRDEV--FVPRTRFQGIQAPKTEEEKKGINVAPMROA 207

```

## RESULT 6

```

US-11-072-512-3814
: Sequence 3814, Application US/11072512
: Publication No. US20060029945A1
: GENERAL INFORMATION:
: APPLICANT: ISOGAI, TAKAO
: APPLICANT: SUGIYAMA, TOMOYASU
: APPLICANT: OTSUKI, TETSUJI
: APPLICANT: WAKAMATSU, AI
: APPLICANT: SATO, HIROYUKI
: APPLICANT: ISHII, SHIZUKO
: APPLICANT: YAMAMOTO, JUN-ICHI
: APPLICANT: ISONO, YUUKO
: APPLICANT: HIO, YURI
: APPLICANT: OTSUKA, KAORU
: APPLICANT: MAGAI, KEIICHI
: APPLICANT: IRIE, RYOTARO
: APPLICANT: TAMECHIKA, ICHIRO
: APPLICANT: SEKI, NAOHICO
: APPLICANT: YOSHIKAWA, TSUTOMU
: APPLICANT: OTSUKA, MOTOTYKI
: APPLICANT: NAGAHARI, KENJI
: APPLICANT: MASUHO, YASUHIKO
: TITLE OF INVENTION: Novel full length cDNA
: FILE REFERENCE: 08435-0191
: CURRENT APPLICATION NUMBER: US/11/072,512
: CURRENT FILING DATE: 2005-03-07
: PRIOR APPLICATION NUMBER: US 60/350,978
: PRIOR FILING DATE: 2002-01-25
: PRIOR APPLICATION NUMBER: JP 2001-379298
: PRIOR FILING DATE: 2001-11-05
: NUMBER OF SEQ ID NOS: 4096
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 3814
: LENGTH: 571
: TYPE: PRT
: ORGANISM: Homo sapiens
US-11-072-512-3814

Query Match      10.7% Score 68; DB 7; length 571;
Best Local Similarity 28.3%; Pred. No. 9.2;
Matches 30; Conservative 15; Mismatches 29; Indels 32; Gaps 7,

QY      29  WGVHO----EAMNLAIMHPKRVGSATTYQYTDKSHDGTARVPDEDTLS--TVSGWLSL 82
      ||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      4  WGEKRGAGESPKLAVYATRKTRS--VRSQEDQWYLG---YPDDQSSGFSYSWYWS 57

QY      83  LGTOS-----PLAD-----ELRAVVRIGWPAAYALGHEHL 112
      ||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      58  VGSSEKGEAGALDQPDQHDVRLDELGLRPAARSGVPGV---BHV 99

RESULT 7
US-11-000-463-745
: Sequence 745, Application US/11000463
: Publication No. US20050266423A1
: GENERAL INFORMATION:
: APPLICANT: Tang, Y Tom
: APPLICANT: Liu, Chenghua
: APPLICANT: Asundi, Vinod
: APPLICANT: Chen, Rui-hong
: APPLICANT: Qian, Xiaohong B.
: APPLICANT: Wang, Zhiwei
: APPLICANT: Wehrman, Tom
: APPLICANT: Zhang, Jie
: APPLICANT: Zhou, Ping
: APPLICANT: Cao, Yi-Cheng
: APPLICANT: Drmanac, Radoje T.
: TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
: FILE REFERENCE: 785CIP4CN
: CURRENT APPLICATION NUMBER: US/11/000,463
: CURRENT FILING DATE: 2004-11-29
: PRIOR APPLICATION NUMBER: 10/291,265
: PRIOR FILING DATE: 2002-11-08

```

```

PCT/US01/02623
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/611,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-09-15
NUMBER OF SEQ ID NOS: 944
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 745
LENGTH: 231
TYPE: PRT
ORGANISM: Homo sapiens
US-11-000-463-745

Query Match      10.4%; Score 66.5; DB 7; Length 231;
Best Local Similarity 24.4%; Pred. No. 4.5;
Matches 21; Conservative 12; Mismatches 34; Indels 19; Gaps 3

Cy      18 AMARCGRIGWGQHGAEAM-----NLAWPRKYGASATIQVDRSHDGRTPAGDEI 72
          ||||| | | | | : :: : || | | : : : : : : : : : : : : : : : :
Db      17 AWPALCRKYKRGRGCVFLTLTTATSYAIV---VWLIWITTYGNKHNSPTWDPLTALI 72
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Oy      73 TSTVSGW-----LSELTGSP 88
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      73 ALANMAAFVLFYIPEVSQYTKRSP 98

RESULT 8
US-11-000-463-273
; Sequence 273, Application US//11000463
; Publication No. US20050266423A1
GENERAL INFORMATION:
APPLICANT: Tang, Y Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Chen, Rui-hong
APPLICANT: Qian, Xiaohong B.
APPLICANT: Wang, Zhiwei
APPLICANT: Wehrman, Tom
APPLICANT: Zhang, Jie
APPLICANT: Zhou, Ping
APPLICANT: Cao, Yi-Cheng
APPLICANT: Drmanac, Radcoje T.
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
FILE REFERENCE: 785CIP4CN
CURRENT APPLICATION NUMBER: US//11/000,463
CURRENT FILING DATE: 2004-11-29
PRIOR APPLICATION NUMBER: 10/291,265
PRIOR FILING DATE: 2002-11-08
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: 09/922,279
PRIOR FILING DATE: 2001-08-03
PRIOR APPLICATION NUMBER: 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: 09/611,746
PRIOR FILING DATE: 2000-07-17
PRIOR APPLICATION NUMBER: 09/631,451
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 09/633,870
PRIOR FILING DATE: 2000-09-15
NUMBER OF SEQ ID NOS: 944
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 273
LENGTH: 466
TYPE: PRT
ORGANISM: Homo sapiens
```





Best Local Similarity 28.2%; Pred. No. 4;  
Matches 24; Conservative 7; Mismatches 29; Indels 25; Gaps 4;

Qy 38 NLAIW---HPRKQSAITVOTDR-----SHDGRARVGDITVSGMLSEIG 84  
Db 34 NYVSWYQGHFGARPKLLITIDVSPRSGVSNRFGSGNTASL-----TISGLIAE-- 84

Qy 85 TOSPLADELARAVRIGDWPAPAYAG 109  
Db 85 ---DEADYCGSRVTSVRSYFVG 106

## RESULT 12

US-11-096-568A-18440  
; Sequence 18440, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nikolai et al.

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

; FILE REFERENCE: 2750-1592PUS2

; CURRENT APPLICATION NUMBER: US/11/096,568A

; CURRENT FILING DATE: 2005-04-01

; NUMBER OF SEQ ID NOS: 34471

; SEQ ID NO 18440

; LENGTH: 408

; TYPE: PRT

; ORGANISM: Zea mays subsp. mays

; NAME/KEY: misc\_feature

; LOCATION: (1)..(408)

; OTHER INFORMATION: Ceres Seq. ID no. 12365194

US-11-096-568A-18440

Query Match 10.0%; Score 63.5; DB 7; Length 408;  
Best Local Similarity 47.5%; Pred. No. 20;  
Matches 19; Conservative 3; Mismatches 9; Indels 9; Gaps 4;

Qy 12 RAVGPYAMAGRCGRIGR---WGVHOEAMNMLAIWHPKV 47  
Db 263 RAAGA---VGRCCR-GRLAVAGGVHQPRL-LPRMHPKV 297

## RESULT 13

US-11-096-568A-18439  
; Sequence 18439, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nikolai et al.

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

; FILE REFERENCE: 2750-1592PUS2

; CURRENT APPLICATION NUMBER: US/11/096,568A

; CURRENT FILING DATE: 2005-04-01

; NUMBER OF SEQ ID NOS: 34471

; SEQ ID NO 18439

; LENGTH: 412

; TYPE: PRT

; ORGANISM: Zea mays subsp. mays

; NAME/KEY: misc\_feature

; LOCATION: (1)..(412)

; OTHER INFORMATION: Ceres Seq. ID no. 12365193

US-11-096-568A-18439

Query Match 10.0%; Score 63.5; DB 7; Length 412;  
Best Local Similarity 47.5%; Pred. No. 20;  
Matches 19; Conservative 3; Mismatches 9; Indels 9; Gaps 4;

Qy 12 RAVGPYAMAGRCGRIGR---WGVHOEAMNMLAIWHPKV 47  
Db 267 RAAGA---VGRCCR-GRLAVAGGVHQPRL-LPRMHPKV 301

## RESULT 14

US-11-096-568A-18438  
; Sequence 18438, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nikolai et al.

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

; FILE REFERENCE: 2750-1592PUS2

; CURRENT APPLICATION NUMBER: US/11/096,568A

; CURRENT FILING DATE: 2005-04-01

; NUMBER OF SEQ ID NOS: 34471

; SEQ ID NO 18438

; LENGTH: 417

; TYPE: PRT

; ORGANISM: Zea mays subsp. mays

; NAME/KEY: misc\_feature

; LOCATION: (1)..(417)

; OTHER INFORMATION: Ceres Seq. ID no. 12365192

US-11-096-568A-18438

Query Match 10.0%; Score 63.5; DB 7; Length 417;  
Best Local Similarity 47.5%; Pred. No. 20;  
Matches 19; Conservative 3; Mismatches 9; Indels 9; Gaps 4;

Qy 12 RAVGPYAMAGRCGRIGR---WGVHOEAMNMLAIWHPKV 47  
Db 272 RAAGA---VGRCCR-GRLAVAGGVHQPRL-LPRMHPKV 306

## RESULT 15

US-11-096-568A-24503  
; Sequence 24503, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nikolai et al.

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

; FILE REFERENCE: 2750-1592PUS2

; CURRENT APPLICATION NUMBER: US/11/096,568A

; CURRENT FILING DATE: 2005-04-01

; NUMBER OF SEQ ID NOS: 34471

; SEQ ID NO 24503

; LENGTH: 447

; TYPE: PRT

; ORGANISM: Zea mays subsp. mays

; NAME/KEY: misc\_feature

; LOCATION: (1)..(447)

; OTHER INFORMATION: Ceres Seq. ID no. 12435494

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (52)..(52)

; OTHER INFORMATION: Xaa is any aa, unknown or other

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (58)..(58)

; OTHER INFORMATION: Xaa is any aa, unknown or other

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (59)..(59)

; OTHER INFORMATION: Xaa is any aa, unknown or other

US-11-096-568A-24503

Query Match 10.0%; Score 63.5; DB 7; Length 447;  
Best Local Similarity 25.0%; Pred. No. 22;  
Matches 23; Conservative 7; Mismatches 41; Indels 21; Gaps 3;

Qy 11 IRVGPYAMAGRCGRIGR---WGVHOEAMNMLAIWHPKV 70  
Db 4 LRVVG---WMTTAAAPMGVHRCRCGTPLARPPAATSSU-----GAGQ 45

Qy 71 BITSTVSGWLSL--GTOSPLADELARAARI 99  
| | | | | : | | | | | :  
Db 46 LRVGTGXWLMDDXGGGGAGADYAREMEVAARV 77

Search completed: March 23, 2006, 06:29:41  
Job time : 3.8797 secs